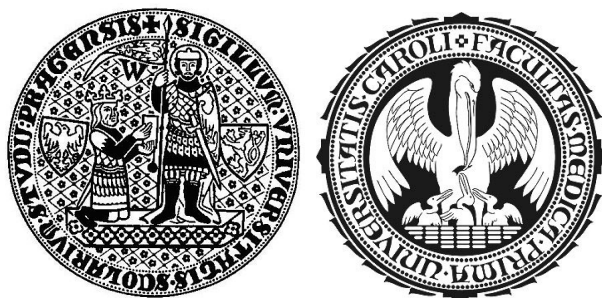


Univerzita Karlova v Praze

1. lékařská fakulta

Obor: Molekulární a buněčná biologie, genetika a virologie

Mgr. Anastasiya Zidkova



DIZERTAČNÍ PRÁCE

**Úloha polymorfních markerů DNA v identifikaci
osob a určování vybraných fenotypových znaků**

**Role of polymorphic DNA markers in personal
identification and determination of selected
phenotypic traits**

Školitel: doc. RNDr. Marie Korabečná, Ph. D.

Konzultant: Ing. Aleš Hořínek

Praha, 2013

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem řádně uvedla a citovala všechny použité prameny a literaturu. Současně prohlašuji, že práce nebyla využita k získání jiného nebo stejného titulu.

Souhlasím s trvalým uložením elektronické verze mé práce v databázi systému meziuniverzitního projektu Theses.cz za účelem soustavné kontroly podobnosti kvalifikačních prací.

V Praze, 30.5.2013

ANASTASSIYA ZIDKOVA

Podpis

Identifikační záznam:

ZIDKOVA, Anastassiya. *ÚLOHA POLYMORFNÍCH MARKERŮ DNA V IDENTIFIKACI OSOB A URČOVÁNÍ VYBRANÝCH FENOTYPOVÝCH ZNAKŮ. [ROLE OF POLYMORPHIC DNA MARKERS IN PERSONAL IDENTIFICATION AND DETERMINATION OF SELECTED PHENOTYPIC TRAITS]*. Praha, 2013. Počet stran 114. Dizertační práce. Univerzita Karlova v Praze, 1. lékařská fakulta, Ústav biologie a lékařské genetiky 1. LF UK a VFN v Praze. Školitel Korabečná, Marie. Konzultant Hořínek, Aleš.

PODĚKOVÁNÍ

Na prvním místě bych chtěla poděkovat své školitelce doc. RNDr. Marii Korabečné, Ph. D. a konzultantovi Ing. Aleši Hořínkovi za odborné vedení, podporu, cenné rady a připomínky a za možnost podílet se na problematice forenzní genetiky. Rovněž bych chtěla poděkovat Ing. Věře Kebrdlové, Mgr. Pavle Coufalové, Mgr. Vlastovi Stenzlovi, RNDr. Pavlu Čapkovi, Ph. D. za jejich odbornou pomoc a spolupráci při řešení dílčích projektů dizertační práce.

Dále bych chtěla poděkovat celému kolektivu laboratoře molekulární genetiky na Ústavu biologie a lékařské genetiky 1. LF UK a VFN v Praze za jejich cenné rady při provádění experimentů. Osobně bych chtěla poděkovat Mgr. Leně Obeidové za kritické poznámky k textu dizertační práce. Ráda bych poděkovala všem dobrovolníkům, kteří se podíleli na studii fenotypových markerů.

V neposlední radě bych chtěla poděkovat své rodině za podporu a pochopení během celé doby studia.

Obsah

SEZNAM ZKRATEK	1
ABSTRAKT	5
ABSTRACT	6
1 ÚVOD	7
1.1 Historie vývoje forenzní genetiky	7
1.2 Struktura lidského genomu	10
1.3 Typy forenzně využitelných markerů DNA	10
1.3.1 STR	11
1.3.2 SNP	15
1.3.3 INDEL	20
1.4 Molekulární podstata pigmentace	21
1.5 Určování fenotypu ve forenzní genetice	26
2 HYPOTÉZY A CÍLE	28
3 MATERIÁL A METODY	29
3.1 Soubor vzorků	29
3.1.1 Soubor vzorků pro studium Investigator DIPplex	29
3.1.2 Soubor vzorků pro studium Mentype [®] Argus X-8	29
3.1.3 Soubor vzorků pro studium Decaplex X-STR	29
3.1.4 Soubor vzorků pro studium Fenotypové SNP	29
3.2 Zkoumané markery	30
3.2.1 Investigator DIPplex	30
3.2.2 Argus X-8	32
3.2.3 Decaplex X-STR	33
3.2.4 Fenotypové SNP markery	33
3.3 Izolace DNA	34
3.4 Genotypizace	35
3.4.1 Investigator DIPplex	35
3.4.2 Fenotypové SNP markery	37
3.5 Statistická analýza	40

3.5.1	Testování Hardy-Weinbergovy rovnováhy	40
3.5.2	Testování vazebné nerovnováhy	42
3.5.3	Určování forenzních parametrů	43
3.5.4	Mezipopulační srovnání	47
3.5.5	Rekonstrukce haplotypů	48
3.5.6	Asociační studie	49
3.5.7	Multifaktorová dimenzionální redukční analýza	50
3.5.8	Analýza pomocí Bayesovských sítí	53
3.5.9	Validace vytvořeného modelu	55
4	VÝSLEDKY	56
4.1	Investigator DIPplex	56
4.1.1	Optimalizace amplifikace a fragmentační analýzy	56
4.1.2	Populační parametry	57
4.1.3	Forenzní parametry	57
4.1.4	Mezipopulační porovnání	60
4.2	Mentype® Argus X-8	62
4.2.1	Populační parametry	62
4.2.2	Forenzní parametry	65
4.2.3	Mezipopulační porovnání	65
4.3	Decaplex X-STR	67
4.3.1	Populační parametry	67
4.3.2	Forenzní parametry	68
4.3.3	Mezipopulační porovnání	68
4.4	Fenotypové SNP	70
4.4.1	Populační parametry	70
4.4.2	Asociační test	72
4.4.3	Analýza MDR	76
4.4.4	Rekonstrukce haplotypů	77
4.4.5	Vytvoření Bayesovy sítě	79
4.4.6	Validace Bayesovy sítě	80
5	DISKUZE	81
5.1	Investigator DIPplex	81

5.2	Mentype [®] Argus X-8	83
5.3	Decaplex X-STR	85
5.4	Fenotypové SNP	86
6	ZÁVĚR	92
	SEZNAM POUŽITÉ LITERATURY	93

SEZNAM ZKRATEK

AIM Ancestry Informative Marker

API Average Paternity Index, průměrný paternitní index

ARA Androgen Receptor

BS Bayesovská síť

cAMP Cyclic Adenosine Monophosphate, cyklický adenosinmonofosfát

CI Confidence intervals

CPI Celkový paternitní index

CPD Combined Power of Discrimination, celková diskriminační síla

CPM Combined Probability of a Match, celková pravděpodobnost shody

CV Cross Validation

ChrX Chromozóm X

DCT Dopachrome Tautomerase

ddNTP Dideoxynucleotide Triphosphate

E Expected, očekávaná hodnota

ELB Excoffier-Laval-Balding

GRCh Genome Reference Consortium Human

HERC2 HECT and RLD domain containing E3 ubiquitin protein ligase 2

HLA (Human Leukocyte Antigen

HPRTB Hypoxanthine Phosphoribosyltransferase

HTLF Helicase-Like Transcription Factor

HWE Hardy-Weinberg Equilibrium

INDEL Inzerčně-Deleční polymorfizmus

IRF4 Interferon Regulatory Factor 4

ISFG International Society for Forensic Genetics

kbp Kilobase pair, kilobáze

KITLG Kit Ligand Protein

KITRT Kit Receptor Tyrosine Kinase

LD Linkage Disequilibrium, vazebná nerovnováha

LEF1 Lymphoid Enhancer Factor-1

LG Linkage Group, vazebná skupina

LINE Long Interspersed Nuclear Elements, dlouhé rozptýlené jaderné elementy

LR Likelihood Ratio, věrohodnostní poměr

LTR Long Terminal Repeat

MAF Minor allele frequency, frekvence minoritní alely

Mbp Megabase pair, megabáze

MCR1 Melanocortin-Receptor 1

MEC Mean Exclusion Chance, průměrná pravděpodobnost vyloučení

MGB Minor Groove Binding

MITF Microphthalmia - Associated Transcription Factor

MSH Melanocyte Stimulating Hormone

mtDNA Mitochondriální DNA

NCBI National Center for Biotechnology Information, Národní centrum biotechnologických informací

NPH Negativní Prediktivní Hodnota

O Observed, pozorovaná hodnota

OCA2 Oculocutaneous Albinism type 2, okulokutánní albinismus II

Oc Het Očekávaná heterozygozita

OR Odds Ratio, poměr podílů pravděpodobností

PBS Phosphate Buffered Saline

PCR Polymerase Chain Reaction, polymerázová řetězová reakce

PD Power of Discrimination, diskriminační síla

PIC Polymorphism Information Content, informační obsah polymorfizmu

PM Probability of a Match, pravděpodobnost shody

Poz Het, He Pozorovaná heterozygozita

PPH Pozitivní Prediktivní Hodnota

RFLP Restriction Fragment-Length Polymorphism, polymorfyzmus délky restričních fragmentů

SINE Short Interspersed Nuclear Elements, krátké rozptýlené jaderné elementy

SLC24A4 Solute Carrier Family 24 Member 4

SLC24A5 Solute Carrier Family 24 Member 5

SLC45A2 Solute Carrier Family 45 Member 2

SNP Single Nucleotide Polymorphism, jednonukleotidový polymorfizmus

SSR Simple Sequence repeats, repetice jednoduchých sekvencí

STR Short Tandem Repeats, krátké tandemové repetice

SWI-SNF SWItch-Sucrose NonFermentable

TPCN2 Two-Pore Segment Channel 2

TYR Tyrosinase, tyrozináza

VNTR Variable Number Tandem Repeats, variabilní počet tandemových repetice

W Pravděpodobnost otcovství

ABSTRAKT

V současné době probíhá intenzivní výzkum použití genetických polymorfizmů pro analýzu degradovaných vzorků, identifikaci a určení příbuznosti. Dalším výzkumným cílem je rozšíření možností forenzní genetiky směrem k determinaci biogeografického původu a s ním spojených fenotypových charakteristik, jako barva očí, vlasů a pleti.

První část předkládané práce se zabývala zpracováním studie založené na vyšetření populačního vzorku z České republiky pomocí sady markerů Investigator DIPlex (QIAGEN, Německo) obsahující 30 autozomálních inzerčně-delečních polymorfizmů. Diskriminační síla (Power of Discrimination - PD), která udává pravděpodobnost, že dvě náhodně vybrané osoby nebudou mít stejný genotyp, pro celou sadu dosáhla hodnoty 99,999999999%. Tato část studie dospěla k výsledku, že uvedená sada markerů je vhodná jako doplňkový panel markerů pro forenzní identifikaci a určování příbuznosti v české populaci.

Druhá část předkládané práce představuje studii na středochovatském populačním vzorku pomocí sady Mentyne Argus X-8 (QIAGEN, Německo) obsahující 8 polymorfizmů krátkých tandemových repetitiv na chromozómu X (X-STR) rozdělených do 4 vazebných skupin. PD celé sady markerů bylo 99,9999% u mužských vzorků a 99,99999999% u ženských. Tato sada markerů může být použita ve středo-chorvatské populaci jako doplňující sada pro určení příbuznosti a identifikaci.

Další část práce byla věnována populační studii se sadou Decaplex X-STR obsahující 10 markerů X-STR, rovnoměrně lokalizovaných podél chromozómu X (kromě páru DXS6809-DXS6789). PD celé sady dosáhla 99,999% u mužů a 99,9999999% u žen. Dle výsledku této studie lze sadu Decaplex X-STR považovat za robustní a vhodnou pro použití ve forenzní genetice a příbuzenské analýze v České republice.

Poslední část předkládané práce se týkala výběru sady markerů pro určování biogeografického původu a barvy očí. Výsledkem populační studie a statistického modelování byl výběr 5 jednonukleotidových polymorfizmů: rs16891982, rs1426654, rs7495174, rs12913832, rs916977. Přesnost predikce barvy očí této sady SNP dosáhla 98,4%, biogeografický původ byl určen správně ve 100% případech.

Klíčová slova: INDEL, X-STR, SNP, forenzní genetika, fenotypové znaky, populační studie

ABSTRACT

Nowadays intensive research is conducted for application of genetic polymorphisms for degraded samples analysis, identification and kinship determination. Another area of research in forensic genetics is biogeographical and phenotypic traits (eye, hair and skin color) determination.

First part of presented work dealt with population study on the Czech population using Investigator DIPplex (QIAGEN, Germany) marker set containing 30 autosomal insertion-deletion polymorphisms. Power of Discrimination (PD), which is the probability of random selection of two persons with different genotypes, was 99.999999999% for the whole marker set. This part of study concluded that analyzed marker set is suitable as an additional marker panel for identification and kinship determination in the Czech Republic.

Second part of the presented study was devoted to population research of Central Croatia using Mentype Argus X-8 kit (QIAGEN, Germany) containing 8 short tandem repeat polymorphisms located on X chromosomes (X-STR) divided into 4 linkage groups. PD for the whole kit reached 99.9999% and 99.99999999% for males and females, respectively. This kit could be used in Central Croatian population for kinship analysis and for identification as an additional marker panel.

The next part of the presented study was the population research on the Czech population using Decaplex X-STR kit containing 10 X-STR evenly spread across X chromosome (except for DXS6809-DXS6789 pair). PD for the whole marker set reached 99.999% for males and 99.9999999% for females. According to the results of the study, Decaplex X-STR is robust kit suitable for forensic identification and kinship analysis in the Czech Republic.

The last part of the presented study concerns the selection of marker for biogeographical ancestry and eye color determination. The population study and statistical modeling resulted in the selection of 5 single nucleotide polymorphisms: rs16891982, rs1426654, rs7495174, rs12913832, rs916977. Accuracy of the selected model for eye color and biogeographical ancestry was 98.4% and 100%, respectively.

Key words: INDEL, X-STR, SNP, forensic genetics, phenotypic traits, population study

1 ÚVOD

1.1 Historie vývoje forenzní genetiky

Analýza DNA vstoupila do forenzní vědy před více než 25 lety. Zasloužil se o to Alec Jeffreys, který zkoumal dědičnou variaci β -globinového genu u člověka pomocí metody RFLP (Polymorfismus délky restričních fragmentů - Restriction fragment-length polymorphism). Tato metoda ovšem analyzuje jednonukleotidové polymorfismy (Single nucleotide polymorphism - SNP), které mají obecně jenom dvě alely. Avšak po objevu multialelických lokusů sestávajících z variabilního počtu tandemových repetitiv (Variable number tandem repeats - VNTR), začal Alec Jeffreys zkoumat přítomnost těchto variabilních lokusů v lidském genomu. Zjistilo se, že gen pro lidský myoglobin obsahuje tyto variabilní oblasti. Překvapivě se variabilní sekvence DNA, která byla přítomna v myoglobinu, nacházela i v jiných místech lidského genomu. Sonda, jež byla vytvořena pro tento gen, hybridizovala s více lokusy, a tím vytvářela na autoradiografu vzor proužků neboli „otisk DNA“, který se lišil u jednotlivých osob a dle kterého bylo možné posuzovat příbuznost [1].

První případ, u kterého byla použita analýza DNA, se stal v roce 1985 [2]. Týkal se průkazu toho, zda dítě z Ghany, kterému hrozila deportace, patří do rodiny rodáků z Ghany s britským občanstvím. Testy DNA ukázaly, že dítě skutečně do rodiny patří. Řízení proti dítěti bylo odloženo, a mohlo tak zůstat s rodinou v Londýně.

První forenzní případ s použitím důkazů DNA se stal v roce 1986 [3]. Dvě mladé ženy, Lynda Mann a Dawn Ashworth, byly znásilněny a poté zabity ve vesnici v britském hrabství Leicestershire v letech 1983 a 1986. Okolnosti případů, jako místo a způsob provedení naznačovaly, že by se mohlo jednat o stejného pachatele. Pod nátlakem společnosti policie zatkla muže, který se přiznal ke spáchání jedné z vražd, ovšem po provedení testů DNA byl tento muž vyloučen jako zdroj spermatu nalezeného na místech činu. Z tohoto důvodu byl v okolí míst činu zahájen masový odběr DNA dospělých mužů. Bylo otestováno kolem 4000 mužů, ale bezvysledně. Až po roce byl zatčen Colin Pitchfork, neboť bylo zjištěno, že místo něj odevzdal vzorek DNA jiný muž. Vzorek, který byl odebrán od skutečného Colina Pitchforka,

se shodoval se vzorky spermatu nalezenými na místech činů. Colin Pitchfork byl odsouzen k doživotnímu vězení.

I když byly takovéto „otisky DNA“ vytvořené pomocí multilokusových sond silně variabilní, interpretace výsledku takové analýzy byla velmi složitá. Proto byly vytvořeny jednolokusové sondy specifické jenom pro jednu variabilní oblast, jejichž výsledkem byla detekce jednoho či dvou fragmentů u každé osoby [4]. Tato analýza byla používána, dokud nebyla nahrazena polymerázovou řetězovou reakcí (Polymerase Chain Reaction - PCR), která byla objevena v roce 1983 Kary Mullisem [5]. PCR byla poprvé použita ve forenzní genetice pro analýzu genu HLA-DQ α (Human Leukocyte Antigen) pomocí alelově specifických prób [6]. Analýzu HLA vystřídala analýza krátkých tandemových repetitiv (Short Tandem Repeats - STR), které byly v roce 1991 poprvé použity pro identifikaci kosterních nálezů [7]. Poté byla v roce 1993 vytvořena první široce používaná sada markerů STR („quadruplex“), která obsahovala 4 STR [8]. V tomtéž roce byl publikován první případ identifikace obětí hromadného neštěstí pomocí STR [9]. Během tohoto neštěstí, které se stalo ve státě Texas, USA, zemřelo 74 osob. Byla analyzována DNA 61 osoby, a následně byly výsledky genetické analýzy a určování příbuznosti interpretovány.

Společně s autozomálními STR se začaly ve forenzní genetice používat i neautozomální markery. Na začátku devadesátých let dvacátého století byla poprvé použita mitochondriální DNA (mtDNA) pro řešení forenzních případů [10]. Později byly použity STR markery, které byly lokalizovány na chromozómu Y [11]. Jednalo se o případ vraždy, které předcházelo znásilnění oběti. Provedení analýzy markerů Y-STR pomohlo vyloučit neprávem obviněného muže.

Na konci 90. let minulého století se začaly vyvíjet metody pro analýzu vzorků s nízkým počtem kopií DNA. Za takové se považují ty vzorky, jejichž celkový obsah DNA je 200 pg a méně. Právě u vzorků s tak nízkým obsahem DNA může docházet k selhání amplifikace jednotlivých alel anebo lokusů a k dalším stochastickým jevům, které následně ovlivňují kvalitu výsledného profilu DNA a jeho interpretaci. Analýza vzorků s nízkým počtem kopií DNA je spojena s řadou problémů, jako kontaminace a ne vždy robustní výsledky. Přesto tato analýza dovoluje určit genotyp u takzvaných dotykových stop a je také využívána v oblasti lékařské genetiky pro analýzu jednotlivých buněk po mikrodisekci [12].

Poznatky získané v oblasti forenzní genetiky začaly být využívány v oblasti an-

tropologie pro identifikaci nalezených kosterních pozůstatků. Jedním z nejznámějších případů byla identifikace ostatků rodiny posledního ruského cara Mikuláše II. V roce 1991 byl v ruském městě Jekatěrinburg nalezen mělký hrob, ve kterém byl údajně pohřben car Mikuláš II., jeho manželka Alexandra, jejich tři děti, osobní lékař a tři členové služebnictva. Pro analýzu kosterních pozůstatků byly použita kombinace markerů STR a sekvenování mtDNA. Analýza STR potvrdila, že je mezi některými osobami příbuzenský vztah. Analýza mtDNA Mikuláše II. odhalila mitochondriální heteroplasmii, což je přítomnost více mitochondriálních genomů s rozdílnou nukleotidovou sekvencí. Po porovnání sekvencí mtDNA s žijícími příbuznými z matčiny strany bylo zjištěno, že jedna ze sekvencí mtDNA údajné kostry Mikuláše II. se shoduje se dvěma žijícími maternálními příbuznými. Na základě těchto poznatků byl učiněn závěr, že se skutečně jedná o kosterní pozůstatky rodiny Romanovců. [13]. V roce 2007 byly nalezeny kosterní pozůstatky dvou osob, které se nacházely 70 metrů od hrobu s kosterními pozůstatky Romanovců. Dodatečné testování potvrdilo domněnku, že se jedná o děti Mikuláše II. - prince Alexeje a jeho sestru [14].

Téměř okamžitě poté, co se testování DNA začalo používat v soudním řízení jako důkaz, se objevila myšlenka založení databáze obsahující údaje o genotypech jedinců, neboli jejich profily DNA. V roce 1994 byla ve Velké Británii vytvořena první databáze profilů DNA pro autozomální markery, která v roce 2004 obsahovala 2,5 milionů profilů DNA [15]. Následně byla vytvořena mezinárodní databáze profilů pro markery lokalizované na chromozómu Y, která je velmi důležitá pro výpočet pravděpodobnosti náhodné shody a shodného původu porovnávaných vzorků [16]. V České republice byla národní databáze autozomálních profilů DNA založena v roce 2002 a dnes obsahuje několik desítek tisíc profilů z míst činů, odsouzených osob i rodinných příslušníků pohřešovaných osob.

S pokrokem v oblasti forenzní genetiky byl také spojen vznik projektu nevinny (Innocence project - <http://www.innocenceproject.org/>), který se na požádání osob, jež tvrdí, že jsou odsouzeny neprávem, zabývá přezkoumáváním jejich případů. Jedním z prvních odsouzených propuštěných díky analýze DNA po vynesení rozsudku byl Gary Dotson. Tento muž byl odsouzen za únos a znásilnění, ovšem testy HLA-DQ α prokázaly, že nemohl být zdrojem spermatu. Oběť se později přiznala, že měla před únosem pohlavní styk se svým přítelem. Rozsudek pro Garyho Dotsona

byl zrušen v roce 1989, po 8 letech vězení. V současné době byl v USA zrušen rozsudek na základě analýzy DNA u 306 osob. Průměrná doba, kterou tito lidé strávili ve vězení, byla více než 13 let.

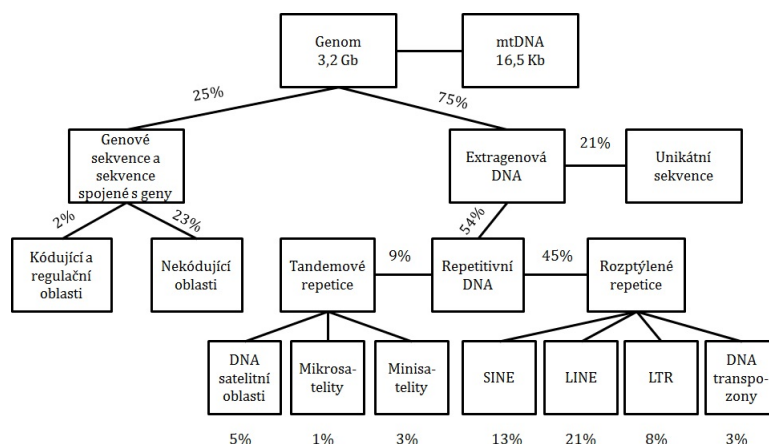
1.2 Struktura lidského genomu

První sekvence lidského genomu byly zveřejněny v roce 2001. Lidský genom se skládá z 3,2 Gb nukleotidů a obsahuje 20000 až 25000 genů. Na obrázku 1.1 je zobrazena struktura lidského genomu. Valnou většinu tvoří nekódující oblasti extragenové DNA, které obsahují pseudogeny, enhancery a silencers, ale také zbytky virových genomů. Kolem 20% genomu tvoří unikátní oblasti s dosud neznámou funkcí [17]. Kódující oblasti představují jenom zhruba 2%. Největší část genomu (50%) je tvořena repetitivními sekvencemi, které jsou buď v blocích neboli tandemech, anebo jsou rozptýlené v genomu. Lidský genom lze rozdělit na několik částí dle jeho variability. Jedna část genomu obsahuje velice konzervativní části, které jsou identické u člověka, myši a krysy. Druhá část je shodná u lidí a ostatních primátů. Třetí je totožná u všech lidí. Poslední část představuje sekvence, které se liší mezi jednotlivými populacemi a uvnitř populace [18]. Forenzní genetika může potenciálně využívat 0,01% lidského genomu, protože takový podíl mají variabilní sekvence, které jsou odlišné u jednotlivých osob. Zhruba 85% těchto variabilních úseků se liší u příslušníků stejné populace [19]. Tyto úseky jsou dobře charakterizovány a využívají se pro identifikaci a určování příbuznosti.

1.3 Typy forenzně využitelných markerů DNA

Ideální vlastnosti, které by měl mít marker používaný ve forenzní genetice pro identifikaci, jsou tyto:

- Vysoká polymorfnost, vysoká schopnost rozlišit osoby mezi sebou
- Jednoduchá, rychlá a levná analýza
- Vytvoření výsledků, které lze jednoduše a jednoznačně interpretovat a které lze jednoduše porovnat mezi laboratořemi
- Nepřítomnost selekčního tlaku na marker



Obrázek 1.1: Rozdělení genomu dle struktury a funkce sekvencí DNA

mtDNA - *mitochondriální DNA*, *SINE* - *short interspersed nuclear elements* = *krátké rozptýlené jaderné elementy*, *LINE* - *long interspersed nuclear elements* = *dlouhé rozptýlené jaderné elementy*, *LTR* - *long terminal repeats*, *dlouhé terminální repetice*

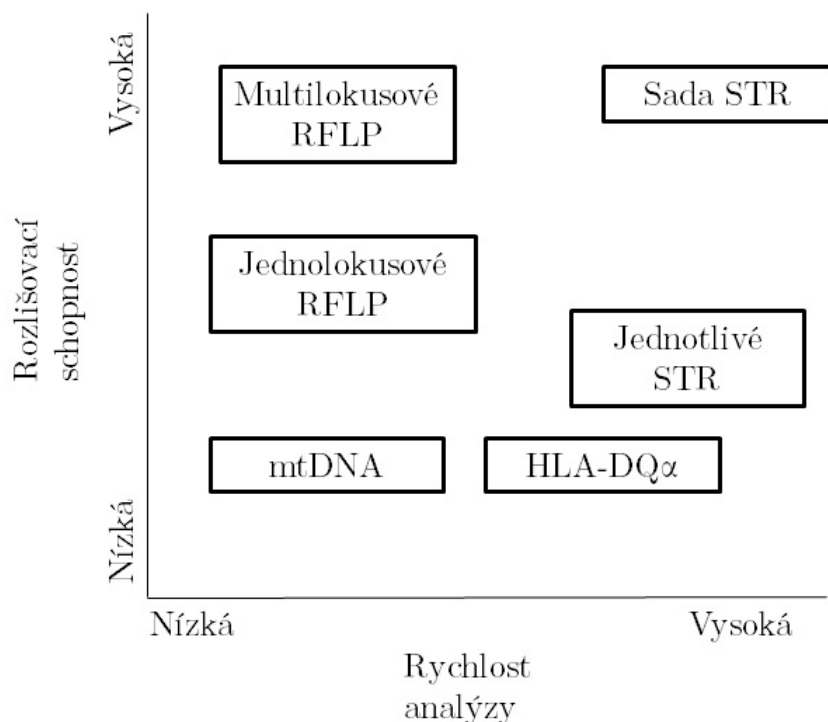
Zdroj: [20]

- Nízká mutační rychlost
- Nepřítomnost selekce

Markery, které se používají nebo používaly ve forenzní genetice, jsou na obrázku 1.2 seřazené dle těchto vlastností. Velmi důležitý je také etický faktor. Zkoumaný marker nesmí prozradit nic o zdravotním stavu testované osoby. Proto se většina v současnosti používaných markerů nachází v extragenových oblastech. Polymorfizmy lze rozdělit do dvou skupin: sekvenční a délkové. Sekvenční polymorfizmy mají stejnou délku, liší se posloupností nukleotidů. Příkladem sekvenčního polymorfizmu je jednonukleotidový polymorfizmus (single nucleotide polymorphism - SNP). Délkové polymorfizmy se liší svou velikostí, příkladem jsou již zmíněné markery STR.

1.3.1 STR

Polymorfizmy STR jsou v současné době nejpoužívanějšími markery ve forenzní praxi. Tento druh polymorfizmů byl nalezen na všech 22 autozomech a gonozomech X, Y. STR se také nazývají mikrosatelitní markery nebo repetice jednoduchých sekvencí (simple sequence repeats - SSR). Pojem „satelitní oblasti“ popisuje úseky



Obrázek 1.2: Srovnání různých identifikačních forenzních markerů

Zdroj: [21]

blízko centromery. Tento pojem vznikl během pokusů využívajících hustotní gradientovou centrifugaci, po níž vznikaly tzv. „satelitní proužky“. Základní repetitivní jednotka se skládá ze 2 až 7 nukleotidů. Obvyklá velikost amplifikovaných fragmentů markerů STR je 100-400 bp [3]. Alely takovýchto markerů se liší počtem repetice, a proto je lze jednoduše rozlišit pomocí kapilární elektroforézy. Pro zviditelnění alel markerů STR se používají fluorescenčně značené primery, s nimiž probíhá amplifikace vybraných úseků DNA. Velkou výhodou markerů STR je možnost vytvoření „multiplexu“, neboli amplifikace více úseků DNA v jedné reakci. Běžné vstupní množství DNA do reakce PCR je kolem 0,5-1 ng. Použití markerů STR je nejvhodnější v případě vzorků obsahujících směs DNA několika osob, protože dovoluje relativně jednoduše rozpoznat, směs kolika osob zkoumaný vzorek obsahuje. Z důvodu zaměření této dizertační práce se soustředíme na markery STR, které jsou lokalizovány na chromozómu X.

Chromozóm X (ChrX) byl pojmenován „X“ jako neznámý, protože již dříve bylo vědcům patrné, že se v mnoha aspektech liší od autozómů [22]. Chromozóm Y byl poté pojmenován zřejmě jako následující písmeno v abecedě [21]. ChrX má kolem 1,5 Mb a obsahuje zhruba 1000 genů, z nichž 99 kóduje proteiny [23]. Většina

úseků ChrX je u mužů hemizigotních. U žen je aktivní jenom jeden chromozóm X, druhý je inaktivován pomocí procesu lyonizace [24] a vytváří Barrovo tělísko. To vysvětluje, proč jsou monozomie a trizomie ChrX dobře tolerovány. Například Turnerův syndrom, který je způsoben monozomií X, je jediná monozomie u lidského organismu, která je slučitelná se životem. Pravděpodobnost monozomie se u ženy s podezřením na Turnerův syndrom zvyšuje, pokud byly pozorovány homozygotní genotypy u několika vázaných markerů na ChrX. Klinefelterův syndrom (nejčastěji 47, XXY) bývá diagnostikován v poměrně pozdním věku a jeho incidence je 1:500 mužů. Tento syndrom může být detekován během určování příbuznosti pomocí X-chromozomálních markerů, když je odhalena heterozygotita hemizygotních úseků. V takovémto případě je analýza pomocí markerů na chromozómu X nevhodná, protože zkoumaný muž obsahuje dva a v některých případech i více ChrX [25]. Bylo zjištěno, že i na inaktivovaném ChrX probíhá transkripce 10% genů, což znamená, že žena má dvojité množství produktů těchto genů. Proto vznikla hypotéza, že tyto rozdíly v expresi genů mohou hrát roli v rozdílech mezi ženami nebo mezi ženami a muži [22].

ChrX obsahuje nadprůměrné množství repetitivních oblastí (56% oproti průměru 45%) [23]. Markery STR lokalizované na chromozómu X (X-STR) byly objeveny spolu s autozomálními markery, ale začaly se používat mnohem později [26]. Jedněmi z prvních byly markery HPRTB (Hypoxanthine Phosphoribosyltransferase) a ARA (Androgen Receptor) [27]. U markeru ARA bylo později zjištěno, že se tato CAG repetice nachází v prvním exonu androgenního receptoru a kóduje polyglutaminový úsek. Zmnožení repetice na větší počet než 43 způsobuje X-vázanou spinální a bulbární svalovou atrofii [28]. Proto z etických důvodů přestal být marker ARA využíván ve forenzní genetice [29].

Velké množství markerů na ChrX dovoluje používat jak vázané, tak i nezávislé markery. Rozdíl spočívá ve výpočtu frekvencí haplotypů. U nezávislých markerů se frekvence haplotypu rovná součinu frekvencí alel. Příkladem nezávislých markerů je sada Decaplex X-STR [30]. U vázaných markerů musí být frekvence haplotypu určena z populačních dat. Důvodem je omezená možnost rekombinace mezi markery, které se nacházejí v těsné blízkosti. Použití vázaných markerů je velmi vhodné při testování příbuznosti. Příkladem vázaných markerů je sada Mentype® Argus X-8 (QIAGEN, Německo).

X-STR mohou být využity jak pro identifikaci, tak i pro určování příbuznosti. Markery na chromozómu X se nepoužívají běžně pro analýzu stop, protože mají menší rozlišovací schopnost než autozomální markery. Také X-STR nejsou vhodné pro určování přítomnosti mužské DNA ve vzorku obsahujícím ženskou DNA. Důvodem je větší pravděpodobnost toho, že alely muže se budou shodovat s alelami ženy, než je tomu u autozomálních markerů. Ovšem X-STR mohou být velmi užitečné, pokud je potřeba určit přítomnost ženské DNA v mužském vzorku. Příkladem takové situace může být vzorek ženských epitelálních buněk pod nehty podezřelého. Situace, že se mužská a ženská DNA budou shodovat ve všech lokusech, může nastat pouze v tom případě, že bude mít žena homozygotní genotyp ve všech analyzovaných lokusech [25].

Tabulka 1.1: Pravděpodobnost přenosu genetické informace (bez započtení mutací)

Dědičnost	Markery AZ	Markery ChrY	Markery mtDNA	Markery ChrX
Matka → syn	50%	-	100%	100%
Matka → dcera	50%	-	100%	50%
Otec → syn	50%	100%	0%	0%
Otec → dcera	50%	0%	0%	100%
Paternální babička → vnučka	25%	-	0%	100%
Maternální babička → vnučka	25%	-	100%	25%
Paternální dědeček → vnuk	25%	100%	0%	0%

AZ - autozomální markery, ChrY - markery lokalizované na chromozómu Y, mtDNA - mitochondriální DNA, ChrX - markery lokalizované na chromozómu X

Zdroj: [21]

Analýza X-STR může doplnit a v komplexních případech určování příbuznosti dokonce mnohonásobně zvýšit informační hodnotu genetické analýzy. V tabulce 1.1 jsou uvedeny pravděpodobnosti přenosu genetické informace. Z této tabulky je patrné, že informativita a vhodnost genetické analýzy s použitím markerů ChrX úzce souvisí s pohlavím zkoumané osoby a s konkrétním případem určování příbuznosti. Obecně lze říct, že X-STR mohou být použity pro řešení komplexních případů určování příbuznosti zahrnující minimálně jednu ženu. Protože dcera dostává od otce kompletní a jeho jediný ChrX, použití X-STR je velmi vhodné, obzvláště pokud není k dispozici matka dcery. X-STR jsou užitečné i pro deficientní případy, jako určování paternity u ženy, pokud je místo jejího údajného otce k dispozici jeho

matka. Dalším případem je určování úplného sourozenectví, kde autozomální markery obecně nejsou vhodné neboť výsledky analýzy podporují tuto hypotézu oproti hypotéze neúplného sourozenectví nebo nepříbuznosti, jenom pokud mají obě osoby stejný genotyp, jehož pravděpodobnost je u úplných sourozenců 25%. Pokud mají dvě ženy stejného otce, potom tyto ženy od něj dostaly stejný ChrX, což lze zjistit pomocí analýzy X-STR. X-STR jsou velmi užitečné, pokud se jedná o případ incestu. Schmidke et al. [31] uvádí případ, kdy bylo zjišťováno, zda je otcem nenarozeného plodu ženského pohlaví otec budoucí matky, nebo její přítel. Pokud by se jednalo o incest, potom by plod neměl žádnou alelu X-STR, kterou by neměla matka. Protože 12 ze 16 paternálních alel nebylo nalezeno u matky, byl učiněn závěr, že otcem plodu je přítel budoucí matky. X-STR plně odpovídají požadavkům na forenzní markery a jak již bylo popsáno, mají široké spektrum využití.

1.3.2 SNP

Jednonukleotidové polymorfizmy (single nucleotide polymorphism - SNP) jsou široce využívány v kvantitativní a lékařské genetice. Pomocí markerů SNP jsou prováděny studie určující soubor genů odpovědných za zkoumaný kvantitativní znak. Tyto markery hrají také roli v kontrole inbrednosti kmenů. V lékařské genetice jsou markery SNP využívány hlavně k vazebné analýze spojené s predikcí postižení zkoumaných osob s geneticky podmíněnou chorobou. Na tomto druhu polymorfizmů je založena metoda RFLP. Vazebná neboli nepřímá analýza pomáhá i v dnešní době u onemocnění, kde jsou časté unikátní mutace a zatím nebylo zavedeno sekvenování nové generace.

Ve forenzní genetice je relativně rozšířena analýza SNP, ovšem jenom u mtDNA. Pro analýzu jaderné DNA zůstávají STR nejpoužívanějšími markery. V současné době začíná forenzní komunita uvažovat o rozšíření analýzy autozomálních SNP markerů pro identifikaci a příbuzenskou analýzu. Důvodem toho jsou výhody, které analýza SNP přináší. Prvním z nich je malá velikost ampliconu, která může být do 100 bp, což je vhodné pro degradované vzorky [21]. Dalším důvodem je možnost plně automatizované analýzy, protože lze, na rozdíl od markerů STR, zvolit metodu nevyžadující určování velikostí fragmentů. Během detekce markerů SNP na kapilární elektroforéze nevznikají „koktavé“ píky, které při analýze pomocí STR způsobují potíže s interpretací vzorků s nízkým množstvím DNA, kdy dochází k amplifikaci

in vitro ke sklouznutí polymerázy a vytvoření arteficiálních fragmentů. Jedním z hlavních důvodů je to, že pomocí SNP lze určit biogeografický původ, pigmentaci očí, vlasů a pleti. Rozdílné vlastnosti markerů SNP a STR jsou shrnuté v tabulce 1.2.

Tabulka 1.2: Rozdílné vlastnosti markerů STR a SNP

Vlastnosti	STR	SNP
Výskyt v genomu	≈ 1 na každých 15 kbp	≈ 1 na každý kbp
Obecná informativita	Vysoká	Nízká, 20-30% informativity STR
Mutační rychlost	≈ 1 na 1000	≈ 1 na 100 000 000
Druh markerů	Di-, tri-, tetra-, penta-nukleotidové repetice s velkým počtem alel	Nejčastěji bialelické markery: A/G, C/T, A/T, C/G, T/G, A/C
Počet alel na marker	Obvykle 5-20	Obvykle 2
Možnost multiplexu	> 10 markerů s různými fluorescentními barvami	Amplifikace do 50 SNP v 1 reakci detekce tisíc SNP pomocí mikročipů
Velikost amplikonů	≈ od 75 do 400 bp	Může být menší než 100 bp
Určení biogeografického původu	Ohraničené	Některé SNP jsou spojeny s biogeografickým původem
Informace o fenotypu	Není	Lze predikovat barvu očí, vlasů a pleti
Hlavní výhoda pro forenzní využití	Velký počet alel dovoluje větší úspěšnost detekce směsných vzorků	Malá velikost amplikonů je vhodná pro degradované vzorky, nízká mutační rychlost může pomoci při určování příbuznosti
Hlavní nevýhoda pro forenzní využití	Interpretace výsledků musí brát v potaz různé artefakty vznikající během analýzy	Nejsou široce využívané základní lokusy, populační substruktura kvůli nízké mutační rychlosti, problematická interpretace směsných vzorků

Zdroj: [21]

Vlastnosti markerů SNP dovolují aplikaci různých metod genotypizace. Zmíníme se o metodách analýzy SNP nejvíce rozšířených v oblasti forenzní genetiky: Sangerovo sekvenování, prodlužování jedné báze a TaqMan assay. Sangerovo sekvenování je velmi rozšířené v lékařské genetice a je založeno na použití modifikovaných nukleotidů - dideoxynukleotidů (ddNTP). Nejprve je provedena amplifikace úseků obsahujících zkoumaný SNP. Potom následuje další amplifikace se směsí jednoho primeru, deoxynukleotidů a fluorescenčně značených ddNTP. Po začlenění ddNTP se zastaví elongace cílového úseku. Tím je vytvořen soubor fragmentů o různé délce podle toho, kdy je inkorporován příslušný ddNTP. Tyto fragmenty jsou následně rozděleny dle velikosti a barvy inkorporovaného ddNTP na kapilární elektroforéze,

jejímž výsledkem je určení posloupnosti nukleotidů ve zkoumaném úseku. Metoda prodlužování jedné báze je podobná Sangerovu sekvenování, proto se také nazývá minisekvenování [21]. První krok této analýzy je stejný jako u předešlé metody, potom ale následuje odstranění neinkorporovaných primerů a nukleotidů. Dalším krokem je přidání speciálních primerů a značených ddNTP. Při této metodě je alelická specifita zaručena komplementaritou na 3' konci primeru. Pokud je tato komplementarita úplná, primer je prodloužen o jeden příslušný ddNTP. Výhodou této metody je možnost tvorby multiplexů, v takovém případě se používá nespecifické přidání polyT na 5' konec primerů. Metoda použití sond TaqMan využívá 5' exonukleázovou aktivitu polymerázy Taq [32]. Reakce obsahuje pár primerů a dvě sondy, z nichž je každá komplementární k jiné alele SNP a má jinou fluorescenční značku. Volná sonda má na sobě zhášec, který brání fluorescenci. Po hybridizaci sondy s templátovou DNA odštěpí polymeráza Taq zhášec, a tím několikanásobně zvýší fluorescenci značky. Tato metoda dovoluje rychlou paralelní analýzu velkého počtu vzorků. Existence více metod pro analýzu markerů SNP dovoluje snadné určení specifity a senzitivity jedné z metod a případně jednoduchou konfirmaci výsledků pomocí jiné metody.

Podle toho, za jakým účelem mohou být markery SNP ve forenzní genetice využity, je můžeme rozdělit do čtyř skupin [33]:

- SNP pro identifikaci: vyžadují vysokou heterozygotitu a nízký koeficient inbreedingu
- SNP pro určování příbuznosti: vazebné skupiny SNP, které představují haplotypové markery pro určování příslušnosti k rodokmenové linii
- SNP pro určování biogeografického původu: vyžadují nízkou heterozygotitu a vysoký koeficient inbreedingu mezi populacemi
- SNP pro určování fenotypových znaků: markery, které poskytují informaci o pravděpodobnosti, se kterou bude zkoumaná osoba mít určitý fenotypový znak

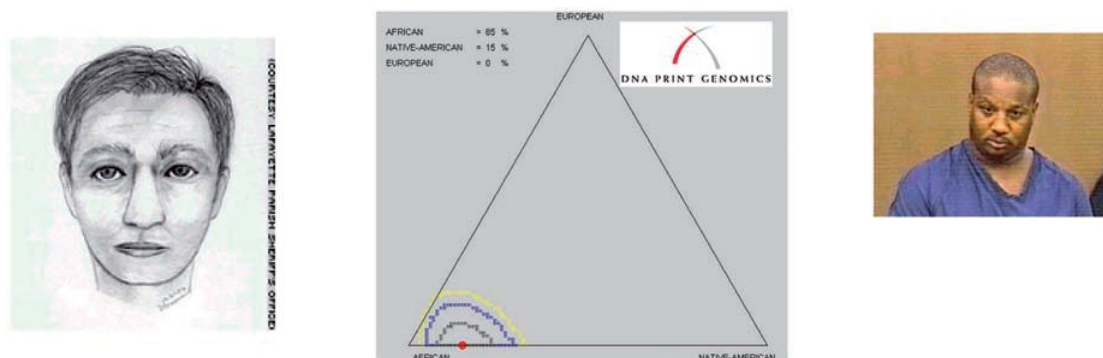
První skupina představuje SNP určené pro identifikaci osob. V této oblasti bylo již realizováno několik velkých projektů. Jedním z nich je již několik let existující konzorcium SNP for ID, které vyvinulo sadu 52 SNP pro identifikaci [34], [35]. SNP obsažené v této sadě splňovaly následující kritéria [34]: 1) výsledný ampikon

nesmí být delší než 120 bp, 2) minimální zjištěná 30% heterozygozita v jedné populaci a minimálně 20% heterozygozita zjištěná u třech populací, 3) lokalizace na od sebe vzdálených úsecích ramének p a q autozomů, 4) minimální vzdálenost 100 kbp mezi zkoumaným markerem SNP a sousedícím genem, 5) absence vazby s markerem STR, který je využíván ve forenzní genetice a 6) absence zjištěných polymorfizmů v oblasti navázání primeru. Pravděpodobnost náhodné shody dosáhla u této sady hodnoty 5×10^{-21} pro bělošskou populaci, což je vyhovující pro forenzní identifikaci, průměrný paternitní index byl 549000. V současné době byla tato sada validována pro populaci Kolumbie [36], Portugalska [37] a Slovinska [38]. Kidd et al. [39] se řídil při výběru identifikačních SNP hlavně mírou mezipopulačního koeficientu inbreedingu. Z analyzovaných 193 markerů bylo vybráno 19 SNP. Všechny vybrané markery byly v Hardy-Weinbergově rovnováze, nebyla mezi nimi vazebná nerovnováha, vzdálenost mezi těmito markery byla 1 Mbp, průměrná heterozygozita byla nejméně 45% pro minimálně tři populace a hodnota koeficientu inbreedingu po porovnání 40 populací byla méně než 0,06. Tyto markery se ukázaly jako velmi robustní a vhodné pro použití nezávisle na původu zkoumané osoby. Ve snaze vyřešit problematické zjišťování směsných vzorků při analýze SNP, byla v průběhu studie Westen et al. [40] vyvinuta sada obsahující soubor trialelických SNP.

Skupiny vázaných markerů SNP přináší nový pohled na určování příbuzností. Studie Ge et al. [41] se zabývala vytypováním vhodných haploskupin. Původně tato skupina vybrala 254 haplobloků na 22 autozomech dle následujících kritérií vztahujících se na jednotlivé haplobloky: minimální počet SNP - 3, mezipopulační koeficient inbreedingu - 0,06, minimální heterozygozita - 20% a minimální počet haplotypů - 3. Po následném zpřísnění kritérií (Hardy-Weinbergova rovnováha a absence vazebné nerovnováhy s ostatními haplobloky) zbylo 24 haplobloků. Souhrnná rozlišovací síla pro tyto haplobloky byla 10^{-12} . V ideálním případě by hodnota informativity jednoho haplobloku měla být podobná jednomu markeru STR. Takový přístup by mohl zvýšit užitečnost relativně malého počtu markerů SNP pro příbuzenskou analýzu.

Výzkum markerů SNP pro určování biogeografického původu (Ancestry Informative Markers - AIM) zkoumá alely, které dostala osoba od obou rodičů, na rozdíl od markerů lokalizovaných na chromozómu Y nebo mtDNA, které dovolují určit příbuznost jen k mužským osobám z paternální linie nebo k ženským osobám z

maternální linie. Autozomální STR nejsou vhodné pro určování původu, protože frekvence jejich alel je v různých populacích obdobná. Další důvod je etický, STR používané pro identifikaci by neměly určovat původ zkoumané osoby. Markery AIM jsou distribuovány po celém světě a mají různou frekvenci v různých populacích. První případ použití markerů AIM u soudu se týkal sériového vraha z Louisiany, kde bylo znásilněno a následně zavražděno pět žen světlé pleti [42]. Byla provedena analýza pomocí STR, která odhalila profil DNA obsahující velmi vzácné alely, který se neshodoval s žádným profilem v databázi. Očití svědci popsali pachatele jako bělocha a byl vytvořen portrét pachatele. Ovšem i po roce a půl vyšetřování nemělo žádné výsledky. Proto byly vzorky DNA z míst činů zaslány do laboratoře DNA-Print, kde bylo provedeno určování biogeografického původu pomocí 71 markerů AIM. Výsledek této analýzy ukázal, že zkoumaná osoba by měla být ze 85% subsaharský Afričan a z 15% americký Indián. Tento výsledek otočil směr vyšetřování, a následně byl zadržen Derrick Todd Lee, jehož profil STR se, jak se poté zjistilo, shodoval s profilem vzorků z míst činů (viz Obrázek 1.3).



Obrázek 1.3: Případ sériovního vraha z Louisiany

Vlevo - portrét údajného pachatele, uprostřed výsledek analýzy 71 markerů AIM, vpravo - zadržený Derrick Todd Lee, jehož profil STR se shodoval s profily z míst činu

Zdroj: [42]

I když mohou AIM nepřímo poskytnout informaci o fenotypových znacích, tato informace je vázaná na biogeografický původ. Proto je především u bělošských populací vhodné použití fenotypových SNP. Tyto SNP by mohly pomoci predikovat fenotypové znaky pachatele, jako jsou barva očí, vlasů a pleti. Fenotypové SNP

mohou být také užitečné pro rekonstrukci vzhledu kosterních pozůstatků neznámé osoby. Protože jedním z cílů této dizertační práce bylo vytvoření sady fenotypových SNP, bude tato kategorie SNP detailněji probrána dále.

1.3.3 INDEL

Forenzně-genetický výzkum obrátil relativně nedávno svou pozornost k inzerčně-delečním polymorfizmům (Insertion-Deletion polymorphism - INDEL). INDEL jsou hojně přítomny v lidském genomu, uvádí se jeden INDEL na 7,2 Kbp lidského genomu. Celkový počet INDEL je neznámý, neboť se údaje mezi studiemi značně liší, což může být způsobeno jak detekční metodou, tak i studovanými soubory vzorků. Markery INDEL představují 16-25% všech polymorfizmů v lidském genomu, velká část z nich je lokalizována v promotorových a exonových oblastech genů, nejvíce z nich se nachází v intronových oblastech [43]. Ve své studii Mills et al. [43] určili 5 hlavních skupin INDEL: 1) inzerce nebo delece jednoho nukleotidu, 2) expanze monomerních nukleotidů, 3) inzerce několika nukleotidů s jednotkou repetice 2-15 nukleotidů, 4) inzerce transpozonů, 5) INDEL obsahující náhodné sekvence DNA od 2 do 9989 nukleotidů. Přibližně 40% INDEL patří do poslední skupiny, z nichž je více než 99% kratších než 100 bp [43]. Třetí skupina zahrnuje i markery STR, které si lze představit jako multialelické INDEL, jejichž alely jsou tvořeny inzercemi a delecemi tandemových repetice [44]. Ovšem v této části se soustředíme na největší skupinu obsahující delece nebo inzerce náhodné sekvence DNA.

INDEL jsou, stejně jako většina markerů SNP, bialelické neboli mají dvě možné alely. Tyto alely mohou být pojmenovány jako „dlouhá“, která má inzerci fragmentu, a jako „krátká“, která nemá zkoumaný fragment. Studie Webera et al. [44] porovnávala úseky, ve kterých se nacházejí INDEL v lidském genomu, a stejné úseky v genomu šimpanzů. Překvapivým výsledkem byla skutečnost, že většina lidských INDEL nebyla polymorfní u šimpanzů. Pomocí tohoto porovnání tato studie také zkoumala původ INDELů - delece nebo inzerce. Ukázalo se, že většina analyzovaných markerů vznikla pomocí delece. Závěrem studie bylo, že kolem 2000 zkoumaných polymorfizmů typu INDEL vzniklo až po vyvinutí člověka/šimpanze/gorily z linie společných předků.

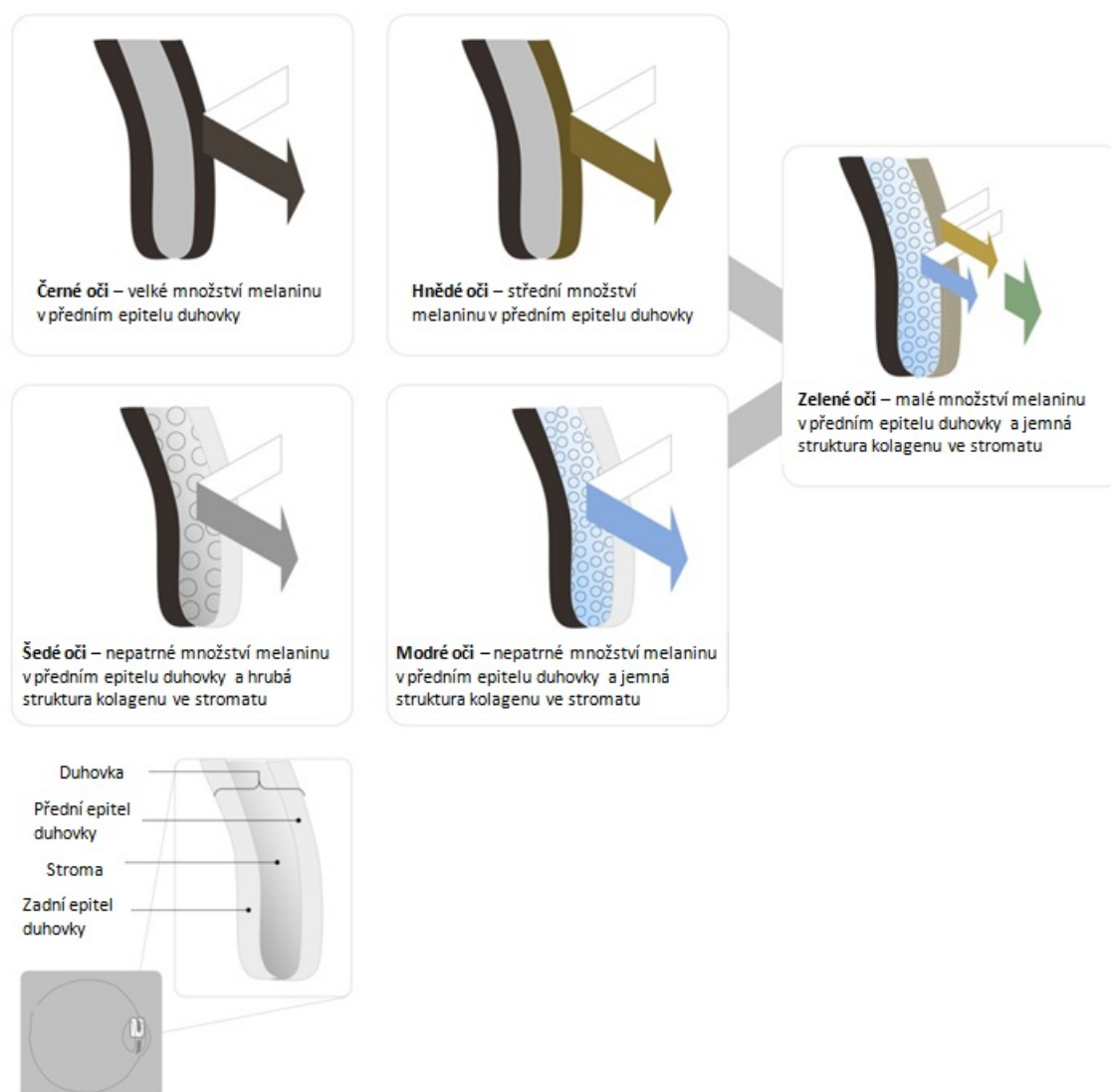
Polymorfizmy INDEL jsou výsledkem jediné mutace, a proto mají nižší mutační rychlost ve srovnání s markery STR. Některé INDEL vznikají inzercí mobilních ele-

mentů. Příkladem je vnoření retrotranspozonu Alu do genomu, čímž vzniká inserce o velikosti kolem 300 bp. Po vnoření retrotranspozonu L1 vzniká inserce dlouhá zhruba 6 Kbp [45]. Velikost ampliconů je u forenzně využitelných INDEL srovnatelná s markery SNP a činí 50-150 bp [46]. Tato velikost předurčuje aplikaci INDEL na degradované vzorky, například kosterní pozůstatky. Protože INDEL mají nižší mutační rychlost, jsou velice vhodné v případech, kdy došlo k údajné mutaci v markerech STR [46], [47], [48]. Vhodnost INDEL pro identifikaci potvrzuje zjištění, že většina zkoumaných INDEL má frekvenci „dlouhé“ alely v rozmezí 30-50% [44]. Velkou výhodou markerů INDEL je možnost detekce těchto markerů pomocí kapilární elektroforézy. Postup detekce je obdobný postupu využívanému pro markery STR: amplifikace s fluorescenčně značenými primery a následné rozdělení amplifikovaných fragmentů na kapilární elektroforéze dle barvy a velikostí. Z toho vyplývá možnost tvorby multiplexů a analýza paralelně až 40 markerů [46]. Pereira et al. [47] představila sadu markerů obsahující 38 INDEL, které mají velikost ampliconů do 160 bp, a vyžadují vstupní množství DNA 300 pg. Investigator DIPplex (QIAGEN, Německo) představuje první a dosud jedinou komerční sadu INDEL. Nehledě na veškeré výhody použití analýzy INDEL pro identifikaci a určování příbuznosti, nejsou tyto markery zatím využívány ve forenzních laboratořích. Jedním z důvodů může být existence rozsáhlých databází genetických profilů vytvořených pomocí analýzy polymorfizmů STR. Protože je však tato problematika v současné době intenzivně zkoumanou oblastí forenzní genetiky, zavedení INDEL pro rutinní použití je jenom otázkou času.

1.4 Molekulární podstata pigmentace

Melanogeneze hraje zásadní roli v procesu pigmentace u lidí, neboť barva lidské pleti, vlasů a očí je z velké části určena množstvím a typem přítomného melaninu. Odstíny barvy očí vznikají také kombinací množství melaninu a struktury kolagenových vláken v duhovce (viz Obrázek 1.4). Proto mohou mít osoby s nepatrným množstvím melaninu jak šedé, tak i modré oči. Bylo zjištěno, že u osob se světle hnědýma očima je větší poměr světlého melaninu (feomelaninu) oproti tmavému melaninu (eumelaninu), zatímco u osob s tmavěhnědýma očima je tento poměr obrácený. Studie neodhalily odlišný počet melanocytů u různých barev očí, proto vznikla hypotéza, že škálu odstínů barev očí vytváří množství melaninu, jeho druh,

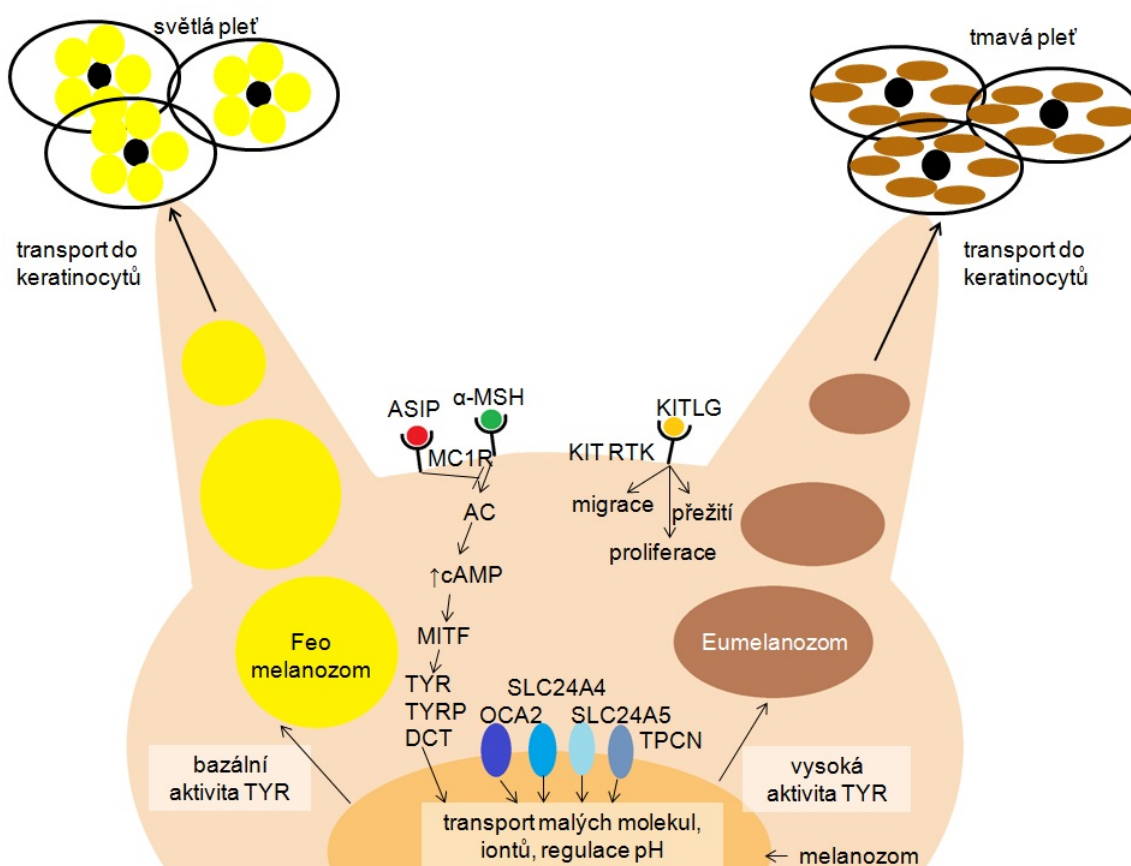
kvalita a hustota [49]. To bylo potvrzeno zjištěním, že se ve světlé pleti kumulují málo zabarvené melanozomy kolem jádra keratinocytů, zatímco v tmavé pleti jsou melanozomy rozmístěny v keratinocytech rovnoměrně. S přibývajícím věkem počet melanocytů klesá, což je patrné u šedivění vlasů. Zároveň se však zvyšuje pigmentace v podobě pigmentových skvrn, což může být vysvětleno proliferací melanocytů, které byly dlouhodobě vystaveny slunci [50].



Obrázek 1.4: Schematické zobrazení vzniku různých typů barvy očí

Zdroj: <http://www.wikimedia.org/>

Melanogeneze probíhá v melanocytech, kde se vytvářejí melanozomy, jež jsou následně transportovány do keratinocytů, kde zajišťují ochranu pokožky proti UV záření. Jeden melanocyt zásobuje zhruba 30 keratinocytů [51]. Na rozdíl od pokožky zůstávají v duhovce melanozomy v melanocytech. Proces melanogeneze v melano-



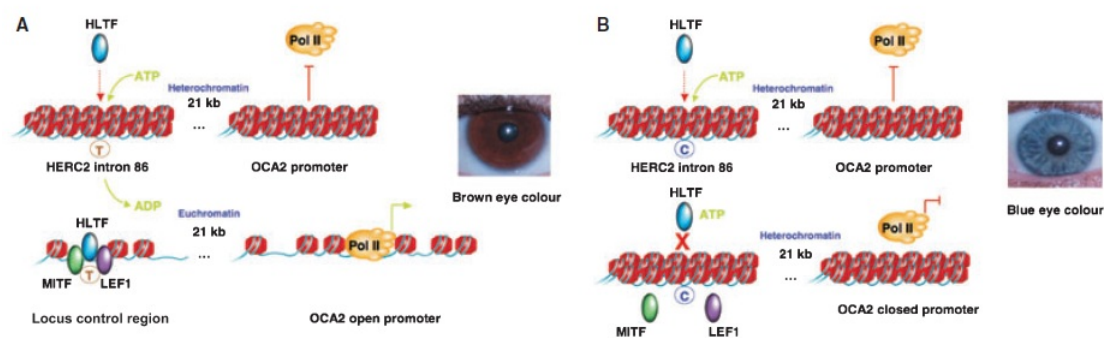
Obrázek 1.5: Schematické zobrazení procesu melanogeneze v melanocytech, popis v textu

cytu je schematicky znázorněn na obrázku 1.5. Tohoto procesu se zúčastňuje kolem 1500 proteinů, zhruba 600 proteinů v každé fázi melanogeneze [52]. Dle průběhu melanogeneze se melanozomy rozdělují do dvou typů: feomelanozomy, u kterých se proces zrání zastaví na prvním stupni, a eumelanozomy, které procházejí úplným procesem zrání. Feomelanozomy produkují světlý (žlutý/červený) feomelanin, zatímco eumelanozomy produkují tmavší (hnědý/černý) eumelanin. Feomelanin je rozpustný, zatímco eumelanin je nerozpustný derivát 3,4 di-hydroxyphenilalaninu (DOPA). Proces melanogeneze začíná produkcí hormonů aktivujících MCR1 (Melanocortin-Receptor 1), které jsou stimulovány UV. Příkladem takového hormonu je α -MSH (α -Melanocyte stimulating Hormone). Oproti tomu ASIP (Agouti Signaling Protein) inhibuje signální dráhu receptoru MCR1 [53]. MCR1 následně zvyšuje v buňce úroveň cyklického adenosinmonofosfátu (Cyclic Adenosine Monophosphate - cAMP), což aktivuje transkripční faktory, které stimulují expresi proteinů zúčastňujících

se zrání melanozomů [51]. Zrání melanozomu začíná vytvořením vezikulů z endoplazmatického retikula. První enzym, který se zúčastňuje přeměny tyrozinu na melanin, je tyrozináza (Tyrosinase - TYR), která je transportována do melanozomu [54]. Tento enzym je přítomen v obou typech melanozomů. Produkce TYR je stimulována transkripčním faktorem MITF (Microphthalmia - Associated Transcription Factor) [55]. V této fázi je zrání feomelanozomů zastaveno a jsou následně transportovány do keratinocytů. Eumelanozomy pokračují v procesu zrání pomocí dalších enzymů, jako například TYRP1 (Tyrosine-Related Protein 1) a DCT (Dopachrome Tautomerase). Tyto proteiny jsou nutné výhradně u tvorby eumelaninu. Rozhodnutí o tom, který druh melaninu bude produkován, závisí na několika faktorech. Hlavním z nich je aktivita enzymu TYR, která je důležitá pro syntézu obou typů melaninu. Pokud má tento enzym bazální aktivitu, potom jsou produkovány jen feomelanozomy. Vysoká aktivita TYR stimuluje enzymy TYRP1 a DCT, což následně vede k produkci eumelaninu. Aktivita TYR závisí na úrovni pH, která je regulována cAMP, a samozřejmě na aktivačních a inaktivačních mutacích samotné TYR [51]. Další skupinou proteinů zúčastňujících se melanogeneze jsou membránové transportéry, které zajišťují transport iontů a malých molekul (jako např. tyrozinu), a regulaci pH, čímž také přispívají k regulaci tvorby melaninu. Například protein P, který je kódován genem pro okulokutánní albinismus II (Oculocutaneous albinism type 2 - OCA2), je transmembranový protein transportující malé molekuly. Iontové kanály, jako například TPCN2 (Two-Pore Segment Channel 2), SLC24A4 (Solute Carrier Family 24 Member 4) a SLC24A5 (Solute Carrier Family 24 Member 5), regulují koncentraci vápníku v melanozomech [54]. Jiný transportér - SLC45A2 (Solute Carrier Family 45 Member 2) hraje pravděpodobně roli v transportu enzymů během maturace melanozomů [56]. Protein KITLG (Kit Ligand Protein) aktivuje KITRT (Kit Receptor Tyrosine Kinase), a tím stimuluje proliferaci a migraci melanocytů. U tohoto proteinu byla nalezena asociace s barvou vlasů ve skotské a dánské populaci [57]. Několik asociačních studií objevilo, že gen IRF4 (Interferon Regulatory Factor 4) lokalizovaný na 6p25.3 je spojen s pigmentačními znaky, ovšem funkce tohoto genu zatím není jasná [58], [59].

Velmi zajímavý je polymorfismus rs12913832, který se nachází v 86. exonu genu HERC2 (viz obrázek 1.6). Tuto heterochromatinovou oblast rozpoznává molekula HTLF (Helicase-Like Transcription Factor) jako součást chromatin-remodelujícího

komplexu SWI-SNF (SWItch-Sucose Non Fermentable). Alela T polymorfizmu rs12913832 (v této studii označována jako A) je rozpoznána DNA-vazebnou doménou HTLF, což způsobuje rozvolnění oblasti dovolující navázání transkripčních faktorů MITF (Microphthalmia - Associated Transcription Factor) a LEF1 (Lymphoid Enhancer Factor-1) a následnou transkripci genu OCA2, který se nachází ve vzdálenosti 21 kbp. Protein OCA2 napomáhá zrání melanozomů, a tím zvyšuje množství melaninu, které má za následek hnědou barvu očí. Pokud je však v tomto polymorfizmu alela C (v této studii označována jako G), potom HTLF nerozpoznává tuto oblast, a ta tak zůstává heterochromatinovou a nedochází k transkripci genu OCA2. To způsobuje nezralost melanozomů u homozygotů pro alelu C (G) a výslednou modrou barvu očí [49].



Obrázek 1.6: Model působení polymorfizmu rs12913832 na determinaci barvy očí

Zdroj: [49]

Mutace způsobující modrou barvu očí nejpravděpodobněji vznikly v severozápadní oblasti Černého moře, odkud před 6000-10000 lety začala migrace do severních částí Evropy. Velká frekvence modrookých lidí v oblastech Skandinávie a Baltského moře naznačuje, že na tento fenotyp mohla působit pozitivní selekce [60]. Vysvětlením může být větší schopnost modrookých lidí ve srovnání s tmavookými zvládat sezonní afektivní poruchu, jedno z hlavních depresivních onemocnění [61].

Polymorfizmy ve všech genech, které se zúčastňují procesu pigmentace, mohou být využity k predikci fenotypových znaků a s nimi spojeného biogeografického původu.

1.5 Určování fenotypu ve forenzní genetice

Molekulárně genetická analýza polymorfizmů DNA má téměř všechny vlastnosti „ideálního identifikačního systému“: důkazní materiál se nemění v čase, technika dovoluje určit jedinečné vlastnosti osoby, pomocí této techniky lze spojit osoby s místem činu a je poměrně levná a rychlá. Přestože současná analýza polymorfizmů DNA produkuje obvykle velice relevantní důkazy, má nízkou diskriminační sílu pro vyšetřovatele. To znamená, že pokud je k dispozici DNA profil podezřelého, který byl zatčen z důvodů nesouvisejících s DNA, anebo po prohledání databáze, analýza shody DNA profilu z místa činu a DNA podezřelého může být velmi silným důkazem. Ovšem pokud nemáme podezřelou osobu, analýza DNA z místa činu pomocí současných metod není schopná určit viditelné fenotypové znaky, které by mohly zmenšit množinu možných pachatelů. Tato skutečnost byla podnětem ke vzniku studií, které se zabývají výzkumem oblastí lidského genomu, jejichž variabilita koreluje s viditelnými fenotypovými znaky.

K rozšíření poznatků o melanogenezi přispěly hlavně celogenomové studie zahrnující různé populace. Takovým příkladem je Rotterdamská studie obsahující celkem 7983 osob [62], [63]. Dalším příkladem je celogenomová studie islandské populace, která byla validována na nizozemské populaci [59], [64]. Tyto studie identifikovaly polymorfizmy v několika majoritních genech, jejichž produkty se zúčastňují melanogeneze. Jedním z průkopníků v aplikaci těchto poznatků do forenzní genetiky byla vědecká skupina Frudakise, která se zabývala výzkumem a následnou aplikací markerů určujících hlavně biogeografický původ, ale také souvisejících s barvou očí, vlasů a pleti [42]. Byly provedeny studie, jejichž snahou bylo určení fenotypových znaků a biogeografického původu kosterních pozůstatků [65]. Nejvíce studií bylo zaměřeno na predikci barvy očí. Barva očí se nemění během ročního období a téměř se nemění v průběhu života, na rozdíl od barvy vlasů a pleti. Během sběru dat je rozřazení jedinců do skupin dle barvy očí relativně jednoduché. V tabulce 1.3 jsou uvedeny vyvinuté sady markerů SNP pro predikci barvy očí. Nejrozšířenějším projektem určování fenotypových znaků, konkrétně barvy očí, je Irisplex [66], [67]. Tato sada obsahuje 6 SNP markerů ze 6 různých genů (HERC2, OCA2, SLC24A4, SLC45A2, TYR, IRF4) a byla vyvinuta na vzorcích buněčných linií projektu Human Genome Diversity [68]. Sada markerů navržená skupinou Ruiz et al. [69] je modifikací Irisplexu obohaceného o jeden marker - rs1129038, který je ve vazbě s

rs12913832. Další sadou je 7-plex [70], který je zaměřen na celosvětovou populaci a obsahuje mimo jiné markery SNP specifické pro africkou (rs6119471) a asijskou populaci (rs1545397, rs885479). Další práce byly založeny na analýze jednotlivých populací: polské [71] a slovinské [72].

Z tabulky 1.3 je patrná zajímavá skutečnost, že všechny sady obsahují marker rs12913832, což potvrzuje důležitost tohoto markeru pro predikci barvy očí. Výběr zbývajících markerů se liší. Důvodem může být populace, na níž byla studie provedena. Proto je velmi vhodné provést výzkum na populaci, pro níž se plánuje zavedení určování fenotypických znaků. Pomocí tohoto výzkumu bude možné vybrat soubor nejinformativnějších markerů a tím zvýšit míru predikce u dotyčné populace.

Tabulka 1.3: Sady markerů SNP vyvinuté pro predikci barvy očí

Marker SNP	Irisplex Walsh et al. 2011 [66]	Ruiz et al. 2013 [69]	7-plex Spichenok et al. 2011 [70]	Pospiech et al. 2011 [71] modré vs nemodré oči	Kastelic, Drobnic 2012 [72]
rs12913832	+	+	+	+	+
rs16891982	+	+	+	-	-
rs1393350	+	+	-	-	+
rs12896399	+	+	-	+	-
rs12203592	+	+	+	-	-
rs1800407	+	+	-	-	+
rs1426654	-	-	+	-	-
rs7495174	-	-	-	-	+
rs6119471	-	-	+	-	-
rs885479	-	-	+	-	-
rs1545397	-	-	+	-	-
rs1805008	-	-	-	-	+
rs1129038	-	+	-	-	-
rs916977	-	-	-	-	-

2 HYPOTÉZY A CÍLE

Předložená práce měla následující cíle a hypotézy:

- Provést populační studii pomocí sady markerů Investigator DIPplex obsahující inzerčně-deleční polymorfizmy na populaci České republiky
- Vyhodnotit zjištěné populační charakteristiky sady markerů Investigator DIPplex na české populaci a porovnat je s ostatními populacemi
- Ověřit hypotézu, že sada markerů Investigator DIPplex je vhodná pro forenzní identifikaci a určování příbuznosti v České republice
- Vyhodnotit zjištěné populační charakteristiky sady X-chromozomálních markerů Mentype® Argus X-8 na středo-chorvatské populaci a porovnat je s ostatními populacemi
- Ověřit hypotézu, že sada markerů Mentype® Argus X-8 je vhodná pro forenzní identifikaci a určování příbuznosti ve středním Chorvatsku
- Vyhodnotit zjištěné populační charakteristiky sady X-chromozomálních markerů Decaplex X-STR na středo-chorvatské populaci a porovnat je s ostatními populacemi
- Ověřit hypotézu, že sada markerů Decaplex X-STR je vhodná pro forenzní identifikaci a určování příbuznosti v České republice
- Vytvořit soubor osob, u kterých bude dostupná informace o barvě očí a původu
- Otestovat tento soubor osob pomocí markerů ve vybraných genech, které se účastní procesu pigmentace
- Vybrat nejlepší model obsahující sadu markerů SNP, který může být použit pro predikci barvy očí v České republice
- Validovat vybraný model a určit jeho míru predikce
- Ověřit hypotézu, že použití kombinace markerů SNP souvisejících s barvou očí a markerů SNP biogeografického původu zlepší predikci barvy očí u neznámých vzorků

3 MATERIÁL A METODY

3.1 Soubor vzorků

3.1.1 Soubor vzorků pro studium Investigator DIPplex

Pro analýzu inzerčně-deleční sady markerů Investigator DIPplex (QIAGEN, Německo) byl použit soubor vzorků DNA z 55 nepříbuzných jedinců České populace. Všechny osoby podepsaly před genetickým vyšetřením informovaný souhlas. Pro stanovení forenzních parametrů, které jsou důležité pro určování paternity, bylo vybráno 11 trií (matka - dítě - otec) s paternitou, která byla prokázána pomocí STR markerů.

3.1.2 Soubor vzorků pro studium Mentype[®] Argus X-8

Pro studii sady X STR markerů Mentype[®] Argus X-8 (QIAGEN, Německo) byly použity vzorky od 178 nepříbuzných osob (78 mužů a 99 žen). Osoby pocházely ze středního Chorvatska, z Karlovacké, Záhřebské, Bjelovarsko-bilogorské a Sisacko-moslavinské oblasti. Studie byla schválena etickou komisí Ústavu lékařského výzkumu a profesního zdraví v Záhřebu, Chorvatsko.

3.1.3 Soubor vzorků pro studium Decaplex X-STR

Pro populační studii sady X-STR markerů Decaplex byly použity anonymizované vzorky 431 nepříbuzných osob, z toho 234 mužů a 197 žen. Všechny osoby, které byly vybrány pro tuto studii, pocházely z České Republiky.

3.1.4 Soubor vzorků pro studium Fenotypové SNP

Tato studie obsahovala dvě množiny dat (viz Tabulka 3.1): trénovací (N=131) a testovací (N=128). Vzorky pro trénovací sadu byly získány od dobrovolníků z České Republiky a Kazachstánu, kteří podepsali informovaný souhlas a vyplnili dotazník obsahující otázky ohledně původu a barvy očí. Testovací anonymizované vzorky bělochů byly získány z Kriministického ústavu Praha, asijská data byla získána z projektu HapMap [73] pomocí vývojového prostředí Rstudio [74].

Tabulka 3.1: Popis souborů vzorků

trénovací množina		Testovací množina	
Barva očí			
Světlá	Tmavá	Světlá	Tmavá
47	84	29	99
Biogeografický původ			
Běloch	Asiat	Běloch	Asiat
100	31	47	81

3.2 Zkoumané markery

3.2.1 Investigator DIPplex

Sada Investigator DIPplex obsahuje 30 INDEL umístěných na autozómech (viz Tabulka 3.2). Vybrané polymorfizmy jsou lokalizovány na 19 autozómech, vzdálenost mezi jednotlivými INDEL je minimálně 10 Mbp. Sada markerů Investigator DIPplex také obsahuje úsek genu pro amelogenin pro určení pohlaví. Amelogenin je gen zubní skloviny, který je přítomen jak na X, tak i na Y chromozómu. Na chromozómu X je lokalizace genu Xp22.1-22.3. Na chromozómu Y se amelogenin nachází na krátkém raménku - Yp11.2. Fragment, který se používá pro určení pohlaví, je na chromozómu Y o 6 bází delší, než na chromozómu X. Úsek amelogeninu, který je analyzován pro určení pohlaví, není polymorfní, ovšem je hojně využíván pro sestavování sad markerů používaných ve forenzní genetice. Úskalí určování pohlaví pomocí tohoto markeru tkví ve skutečnosti, že existují populace, u kterých je v tomto úseku na chromozómu Y delece 6 bází. Výsledkem analýzy vzorku muže s takovouto delecí jsou dva fragmenty o stejné délce, které jsou nerozlišitelné na kapilární elektroforéze. Byly doloženy jedinečné případy výskytu této delece u jedinců v Mexiku [75], Indii [76], Číně [77], Japonsku [78] a Itálii [79]. Ovšem frekvence delece v amelogeninu je velmi vzácná v bělošské populaci, a proto patří amelogenin k nejrozšířenějším markerům pro určení pohlaví.

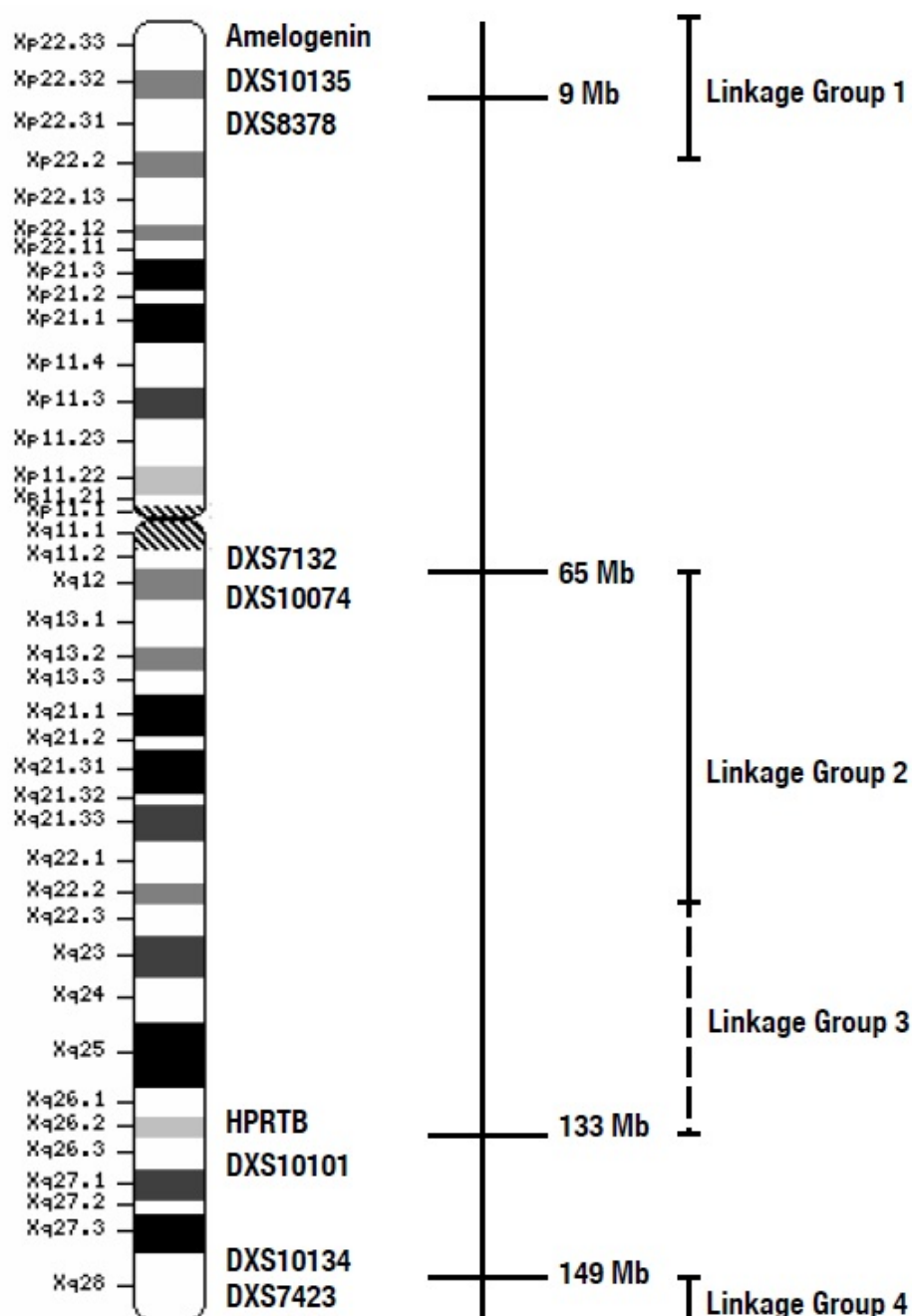
Tabulka 3.2: Seznam polymorfizmů INDEL v sadě Investigator DIPplex

Lokus	Chromozomální lokalizace	Číslo rs	Sekvence polymorfizmu
HLD6	16q13	rs1610905	GCAGGACTGGGCACC
HLD39	1p22.1	rs17878444	CCTAAACAAAAATGGGAT
HLD40	1p32.2	rs2307656	GGGACAGGTGGCCACTAGGAGA
HLD45	2q31.1	rs2307959	CACG
HLD48	2q11.2	rs28369942	GACTT
HLD56	4q25	rs2308292	TAAGT
HLD58	5q14.1	rs1610937	AGGA
HLD64	5q12.3	rs1610935	GACAAA
HLD67	5q33.2	rs1305056	CTACTGAC
HLD70	6q16.1	rs2307652	AGCA
HLD77	7q31.1	rs1611048	TAAG
HLD81	7q21.3	rs17879936	GTAAGCATTGT
HLD83	8p22	rs2308072	AAGG
HLD84	8q24.12	rs3081400	CTTTC
HLD88	9q22.32	rs8190570	CCACAAAGA
HLD92	11q22.2	rs17174476	GTTT
HLD93	12q22	rs2307570	ACTTT
HLD97	13q12.3	rs17238892	AGAGAAAGCTGAAG
HLD99	14q23.1	rs2308163	TGAT
HLD101	15q26.1	rs2307433	GTAG
HLD111	17p11.2	rs1305047	CACA
HLD114	17p13.3	rs2307581	TCCTATTCTACTCTGAAT
HLD118	20p11.1	rs16438	CCCCA
HLD122	21q22.11	rs8178524	GAAGTCTGAGG
HLD124	22q12.3	rs6481	GTGGA
HLD125	22q11.23	rs16388	ATTGCC
HLD128	1q31.3	rs2307924	ATTAAATA
HLD131	7q36.2	rs1611001	TGGGCTTATT
HLD133	3p22.1	rs2067235	CAACCTGGATT
HLD136	22q13.1	rs16363	TGTTT

Zdroj: <http://www.qiagen.com/>

3.2.2 Argus X-8

Sada markerů Mentype® Argus X-8 se skládá ze čtyř vazebných skupin. Každá skupina obsahuje dvojici X-STR markerů, mezi kterými je malá vzdálenost (viz Obrázek 3.1).



Obrázek 3.1: Lokalizace jednotlivých X-STR obsazených v Mentype® Argus X-8

Zdroj: <http://www.biotype.de>

3.2.3 Decaplex X-STR

Na rozdíl od předchozí sady markerů, Decaplex X-STR obsahuje STR markery, které jsou rovnoměrně rozmístěny po celé délce chromozómu X (viz Tabulka 3.3). Decaplex X-STR sdílí s Mentype® Argus X-8 tři společné markery: DXS8378, DXS7132 a DXS7423.

Tabulka 3.3: Umístění markerů ze sady Decaplex X-STR na chromozómu X

Marker	Chromozomální lokalizace	Lokalizace (Mb)*
DXS8378	Xp22.31	9,330
DXS9902	Xp22.2	15,234
DXS7132	centroméra	64,572
DXS9898	Xq21.31	87,682
DXS6809	Xq21.33	94,825
DXS6789	Xq21.33	95,336
DXS7133	Xq22.3	108,928
GATA172D05	Xq23	113,061
GATA31E08	Xq27.1	140,062
DXS7423	Xq28	149,460

**dle NCBI 36*

3.2.4 Fenotypové SNP markery

Tato studie zkoumala 22 markerů SNP, které korelují s biogeografickým původem anebo s barvou očí (viz Tabulka 3.4). SNP byly vybrány z různých literárních zdrojů - jak z výsledků celogenomových asociačních studií, tak i z prací, zabývajících se aplikací určení fenotypu ve forenzní genetice. Při výběru SNP byly preferovány markery nacházející se v kódující oblasti genů. Většina zkoumaných SNPů se nachází na 15. chromozómu v oblasti genů HERC2-OCA2. Bylo prokázáno, že tato oblast má největší rozlišovací schopnost v určování barvy očí ze všech dosud zkoumaných oblastí. V této studii byly zkoumány polymorfizmy i v dalších genech, které se zúčastňují melanogeneze, a nebo se frekvence jejich alel liší u příslušníků různých populací. Z toho důvodu tato studie zahrnovala i markery původu, které jsou mnohdy těsně spojeny s pigmentací.

Tabulka 3.4: Umístění markerů ze sady Decaplex na chromozómu X

Číslo rs	Gen	Cromozomální lokalizace *	Genová oblast	Alely **	Zdroj
rs16891982	SLC45A2 (MATP)	5p13.2	Exon	C/G	[58]
rs12203592	IRF4	6p25.3	Intron	C/T	[58]
rs35264875	TPCN2	11q13.3	Exon	A/T	[59]
rs1393350	TYR	11q14.3	Intron	A/G	[59]
rs12821256	KITLG	12q21.32	Intergenová oblast	C/T	[59]
rs2031526	DCT(TYRP2)	13q32.1	Intron	A/G	[80]
rs12896399	SLC24A4	14q32.12	Intergenová oblast	G/T	[63]
rs1800407	OCA2	15q13.1	Exon	A/G	[81]
rs7495174	OCA2	15q13.1	Intron	A/G	[62]
rs4778138	OCA2	15q13.1	Intron	A/G	[59]
rs1545397	OCA2	15q13.1	Intron	A/T	[82]
rs4778241	OCA2	15q13.1	Intron	A/C	[62]
rs1667394	HERC2	15q13.1	Intron	A/G	[59]
rs12913832	HERC2	15q13.1	Intron	A/G	[83]
rs916977	HERC2	15q13.1	Intron	A/G	[58]
rs1426654	SLC24A5(NCKX5)	15q21.1	Exon	A/G	[84]
rs258322	MCR1	16q24.3	Intron	C/T	[58]
rs1805007	MCR1	16q24.3	Exon	C/T	[59]
rs885479	MCR1	16q24.3	Exon	A/G	[70]
rs4911414	ASIP	20q11.22	Intergenová oblast	G/T	[59]
rs1015362	ASIP	20q11.22	Intergenová oblast	A/G	[59]
rs6119471	ASIP	20q11.22	Intergenová oblast	C/G	[70]

* dle cytogenetických bandů Ensembl (<http://www.ensembl.org/>); ** dle databáze dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>)

3.3 Izolace DNA

DNA byla izolována z bukalních stěrů pomocí Blood Mini Kit (QIAGEN, Německo) dle následujícího postupu.

1. Přidat do 2 ml zkumavky 600 μ l PBS (Phosphate Buffered Saline), 600 μ l

AL pufru a 35 μl Proteinázy K. Do vzniklého roztoku ponořit bukalní stěr a promíchat.

2. Inkubovat 10 minut při 56 °C.
3. Přidat 600 μl 96% etanolu a promíchat.
4. Napipetovat do kolonky 600 μl vzniklého roztoku. Stočit 1 minutu při 6000 \times g. Proces zopakovat 2 krát.
5. Napipetovat do kolonky 600 μl promývacího pufru AW1. Stočit 1 minutu při 6000 \times g.
6. Napipetovat 600 do kolonky μl promývacího pufru AW2. Stočit 3 minuty při 20000 \times g.
7. Napipetovat do kolonky 150 μl elučního pufru AE. Inkubovat 1 minutu při pokojové teplotě, potom stočit 1 minutu při 6000 \times g.
8. Vzniklý eluát přepipetovat do čisté zkumavky.

Po izolaci byla DNA kvantifikována na přístroji Nanophotometer (Implen, USA).

3.4 Genotypizace

3.4.1 Investigator DIPplex

Vzorky DNA byly amplifikovány s použitím značených primerů pro následnou detekci na kapilární elektroforéze.

Pro amplifikaci fragmentů DNA bylo vyzkoušeno několik vstupních množství DNA (0,2 ng, 0,5 ng a 1 ng). Jako nejlepší bylo vyhodnoceno vstupní množství 0,5 ng DNA. PCR reakce byla připravena dle tabulky 3.5 a provedena v cykleru Veriti (Applied Biosystems, USA) v souladu s pokyny výrobce (viz Tabulka 3.6).

Po ukončení amplifikace byla provedena fragmentační analýza. Proto bylo 0,5 μl produktu PCR reakce smícháno s 0,75 μl vnitřního standardu BTO (QIAGEN, Německo) a 12 μl Hi-Di formamidu (Applied Biosystems, USA) a analyzováno na genetickém analyzátoru ABI PRISM 310 (Applied Biosystems, USA) za následujících podmínek:

- Délka kapiláry: 36 cm
- Polymer: POP-4TM(Applied Biosystems, USA)
- Modul: GS STR POP-4 (1 ml) G5
- Doba nástřiku: 5 s
- Napětí při nástřiku: 15kV
- Teplota běhu: 60°C
- Napětí při běhu: 15kV
- Doba běhu: 20 min

Určování genotypů bylo provedeno v programu Genotyper v3.7 (Applied Biosystems, USA) s použitím alelického žebříčku.

Tabulka 3.5: Směs pro PCR reakci Investigator DIPplex

Složka	Objem, μ l
Reakční mix A	2,5
Primery	2,5
Polymeráza Taq	0,3
Voda	6,2
DNA	1,0
Celkový objem	12,5

Tabulka 3.6: Protokol PCR reakce Investigator DIPplex

Teplota	Čas
94 °C	4 min
94 °C	30 s
61 °C	120 s
72 °C	75 s
30 cyklů	
68 °	60 min
10 °	∞

3.4.2 Fenotypové SNP markery

Genotypy byly určeny pomocí metody alelické diskriminace s použitím značených sond TaqMan® MGB™ na přístroji 7900 HT Fast Real-Time PCR System (Applied Biosystems). V každé reakci byla dvojice neznačených primerů a dvě sondy. Každá sonda komplementárně hybridizuje jenom k jedné alele SNP markeru. Molekula sondy má na 5' konci navázanou fluorescenční značku (VIC™ nebo FAM™), na 3' straně je zhášec, který brání emisi fluorescence. K sondě je také připojena molekula, která se váže do malého žlábků DNA (Minor groove binding - MGB), a zvyšuje tím specificitu sondy. Po hybridizaci sondy a během elongační fáze DNA polymeráza pomocí své 5' endonukleázové aktivity odstraní ze sondy fluorescenční značku, čímž se několikanásobně zvýší emise fluorescence (viz Obrázek 3.2). Fluorescence vzorku je měřena před začátkem amplifikace a na konci.

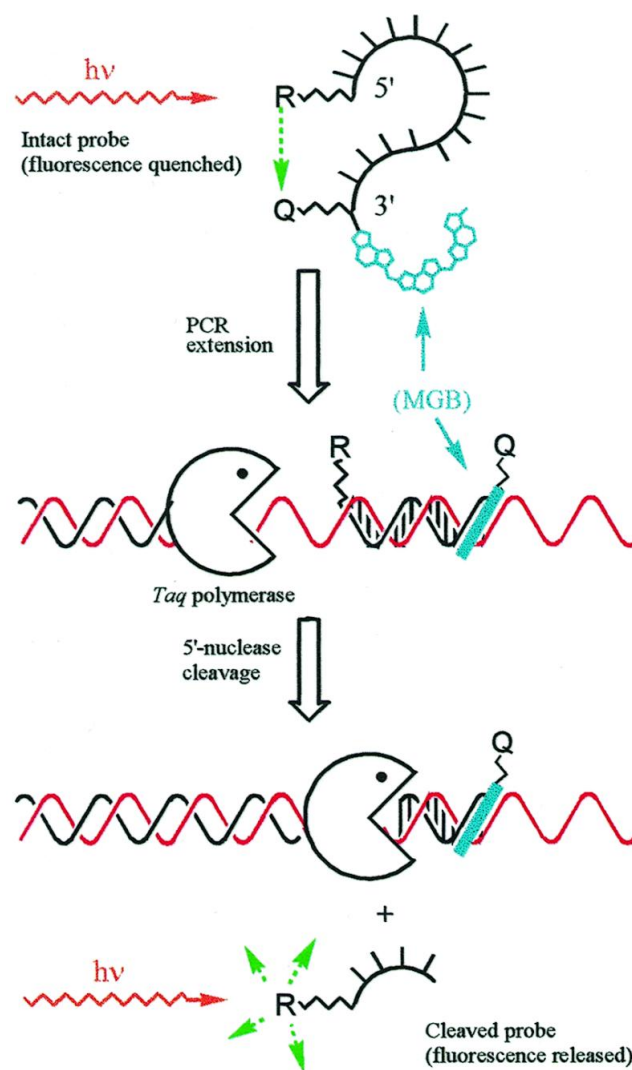
Po změření fluorescence amplifikovaného produktu jsou vzorky vyneseny do grafu (viz Obrázek 3.3). Osa x odpovídá míře fluorescence VIC™, osa y odpovídá míře fluorescence FAM™. Na obrázku představuje každý bod jednu osobu. Homozygoti pro alelu, k níž je komplementární sonda značena VIC™, se nacházejí v pravém dolním rohu. Homozygoti pro alelu, k níž je komplementární sonda značena FAM™, se nacházejí v levém horním rohu. Heterozygoti se nacházejí uprostřed.

Pro analýzu bylo použito 2 μ l neředěného roztoku DNA o koncentraci přibližně 15-20 ng/ μ l. Amplifikace a analýza byla provedena dle pokynů výrobce (viz Tabulky 3.7, 3.8) zvlášť pro každý SNP marker.

Tabulka 3.7: Směs pro PCR reakci fenotypových SNP

Složka	Objem, μ l
TaqMan Universal Mix II, no UNG	5,00
Voda	2,75
Próba a primery	0,25
DNA	2,00
Celkový objem	10,00

Určování genotypů bylo provedeno pomocí programu Sequence Detection System Software v2.4.1 (Applied Biosystems, USA).



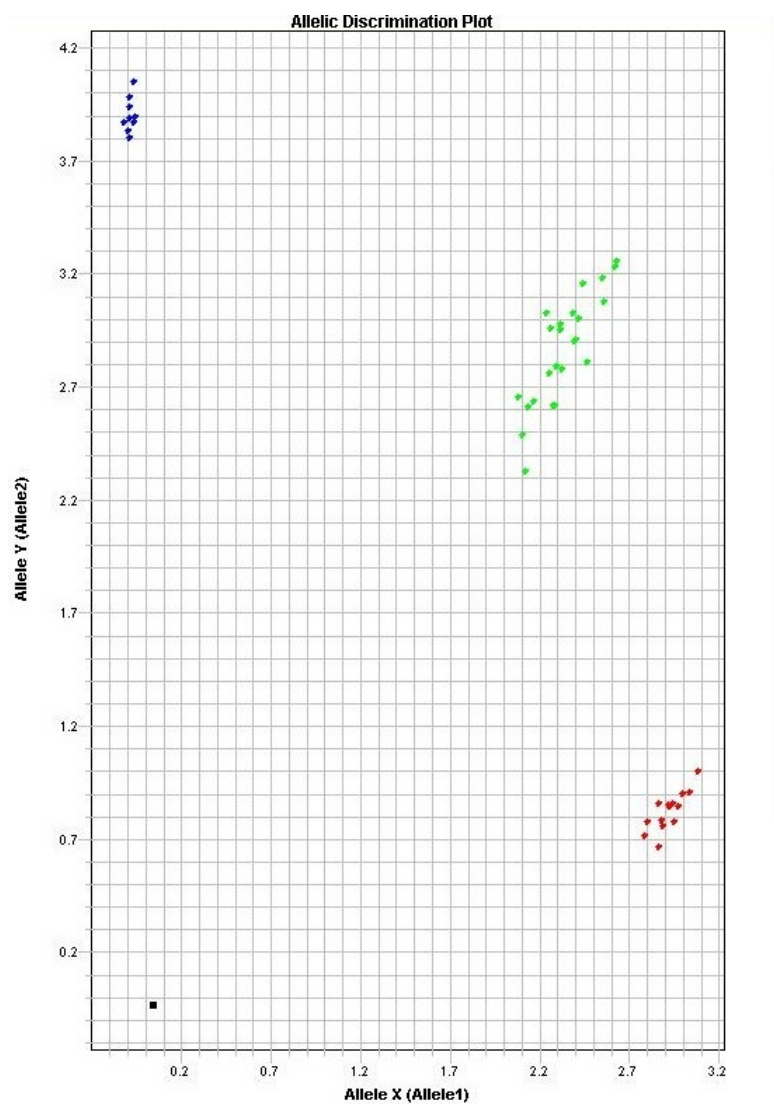
Obrázek 3.2: Princip metody určování genotypu pomocí sond TaqMan[®]MGB[™]

R - fluorescenční značka, *Q* - zhášec, *MGB* - minor groove binding

Zdroj: [85]

Tabulka 3.8: Protokol PCR reakce fenotypových SNP

Teplota	Čas
95 °C	10 min
92 °C	15 s
60 °C	90 s
50 cyklů	



Obrázek 3.3: Výsledek genotypizace pomocí sond TaqMan[®]MGB[™]

3.5 Statistická analýza

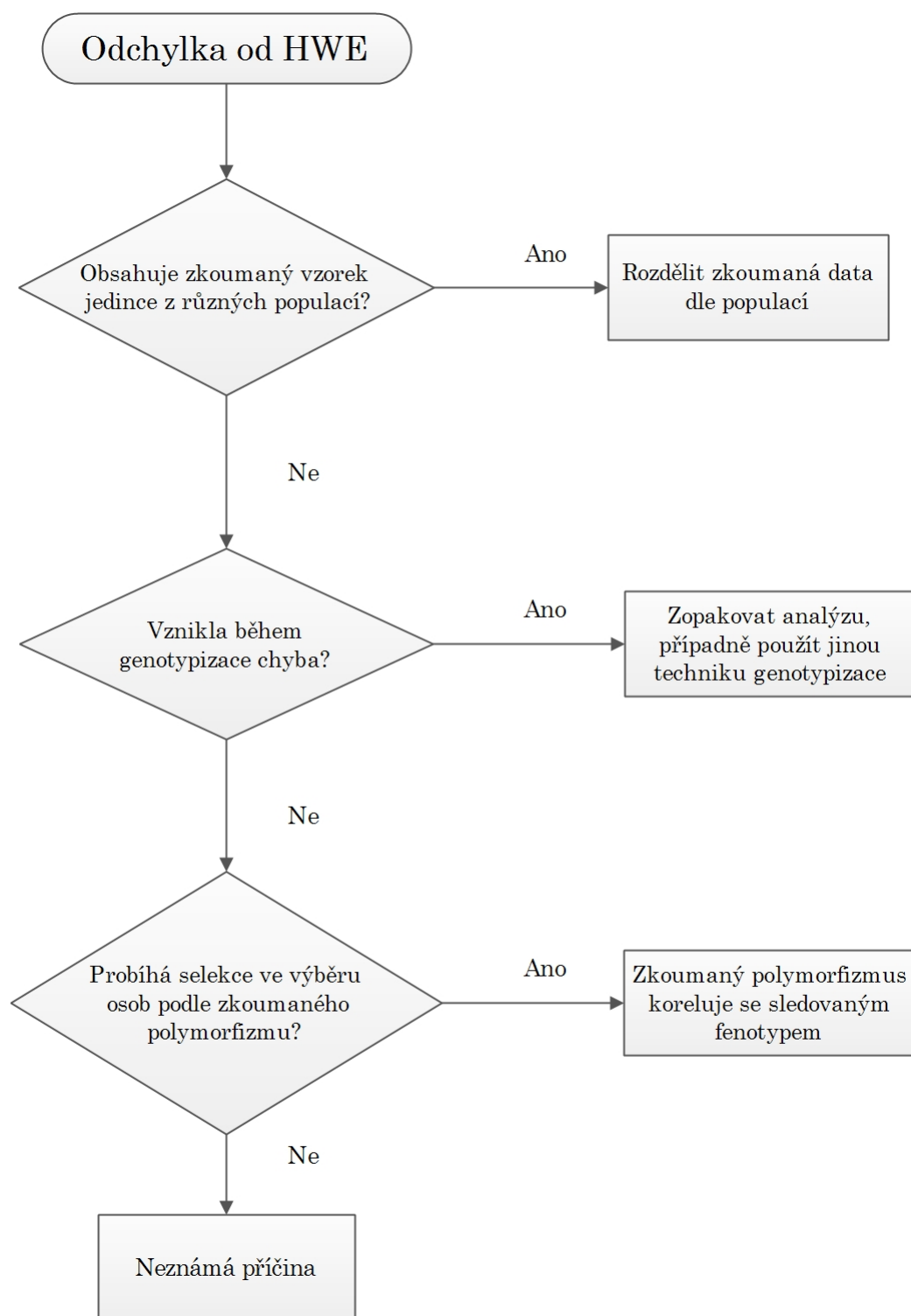
3.5.1 Testování Hardy-Weinbergovy rovnováhy

Testování Hardy-Weinbergovy rovnováhy (Hardy-Wenber equilibrium - HWE) patří k základním statistickým analýzám v populační genetice. Tento test byl měl být proveden před testováním jakýchkoliv hypotéz. Testování HWE dovoluje nejen odhadovat frekvence jednotlivých genotypů, ale i odhalit genotypizační chyby nebo asociaci markerů se zkoumaným fenotypem. Základní předpoklady toho, že populace se nachází v HWE, jsou:

- Náhodné křížení
- Nekonečně velká populace
- Absence mutací
- Absence migrace
- Absence genetického driftu
- Absence selekce

Je zřejmé, že žádná lidská populace nesplňuje všechna tato kritéria. Ovšem ve velkých populacích jsou takové jevy jako migrace, mutace a selekce zanedbatelné, proto se očekává, že tyto populace budou v HWE. U populačního vzorku je potřeba kontrolovat HWE, aby se zjistilo, zda je reprezentativní. Odchylka od HWE může naznačovat buď genotypizační chybu [86], [87], stratifikaci populace, smíšení několika populací do jednoho souboru, anebo naznačovat, že vybraný marker skutečně koreluje se zkoumaným fenotypem [88] (viz Obrázek 3.4).

S odhadem, zda se populace nachází v HWE, je spojen odhad heterozygosity. Tuto veličinu lze vypočítat pomocí odhadu frekvence heterozygotů podle Hardy-Weinbergova zákona. Podíl heterozygotů lze odhadnout pomocí genové diverzity i přes alelické frekvence.



Obrázek 3.4: Vývojový diagram postupu při odhalení odchylky od Hardy-Weinbergovy rovnováhy

V populační studii Investigator DIPplex byl test na HWE vypočítán ručně pomocí χ^2 testu (viz Rovnice 3.1). Pro tuto studii byl také proveden odhad genové diverzity podle práce Chakraborty et al. [89].

$$\chi^2 = \frac{(O - E)^2}{E}, kde \quad (3.1)$$

O - pozorované frekvence genotypů

E - frekvence genotypů odvozené pomocí zjištěných frekvencí alel dle vzorce Hardy-Weinberga pro frekvence genotypů

U vzorků analyzovaných pomocí sady Mentype® Argus X-8 a Decaplex X-STR byl test HWE společně s odhadem genové diverzity proveden pomocí programu Arlequin v3.5 [90]. U vzorků analyzovaných pomocí fenotypových SNP markerů byly zjišťovány odchylky od HWE pomocí programu PLINK v1.07 [91].

Během testování HWE i jiných statistických testů na hodnotu hladiny významnosti byla aplikována Bonferroniho korekce. Důvod korekce spočívá v tom, že pokud testujeme více markerů najednou, existuje pravděpodobnost, že jeden test bude mít signifikantní výsledek jenom díky náhodě. Při aplikaci Bonferroniho korekce se hodnota významnosti vydělí počtem provedených testování. Například, pokud bylo provedeno testování HWE u 30 INDEL markerů ze sady Investigator DIPplex, hodnota významnosti po Bonferroniho korekci je 0,0017. Neexistuje žádné striktní pravidlo ohledně použití Bonferroniho korekce, a je tedy na vědci, zda se rozhodne tuto korekci použít, či nikoliv [92].

3.5.2 Testování vazebné nerovnováhy

Test vazebné nerovnováhy (Linkage Disequilibrium - LD) zkoumá korelaci mezi alelami různých markerů. Hlavní otázka tohoto testu spočívá v určení toho, jak určitá alela jednoho markeru ovlivňuje pravděpodobnost distribuce alel jiných markerů. Jinými slovy tento test zkoumá, zda jsou na sobě analyzované markery závislé či nezávislé. Signifikantní výsledek testu LD může být také způsoben tím, že zkoumaný vzorek není v HWE. Důvodem je skutečnost, že test LD předpokládá frekvence genotypů dle zákona Hardy-Weinberga.

Vazebná nerovnováha byla testována u sady markerů Investigator DIPplex pomocí

programu SNPalyzer v1.2 (Istech, Jižní Korea). U fenotypových SNP markerů, sady Mentype® Argus X-8 a sady Decaplex byl pro testování LD použit program Arlequin v3.5 [90]. U sad obsahujících X-STR markery byl test LD proveden zvlášť pro mužské a ženské vzorky.

3.5.3 Určování forenzních parametrů

Jedním z hlavních kritérií pro marker ve forenzní genetice je jeho informativita. Pokud se jedná o využití markeru za účelem identifikace osoby, zkoumaný marker by měl být co nejpolymorfnější. Jinak řečeno - měl by se v dané populaci vyskytovat v co nejvíce možných genotypech vykazujících přibližně stejnou frekvenci. Další charakteristiky, které hrají roli ve výběru markeru, jsou nízká mutační rychlost a umístění mimo kódující oblast. Poslední charakteristika je důležitá ze dvou důvodů. První je etický, protože zkoumaný polymorfismus nesmí poskytnout informaci o zdravotním stavu osoby, jejíž DNA je analyzována. Druhý důvod je populačně-genetický, protože polymorfismus nacházející se v kódující oblasti by mohl podlehnout selekčnímu tlaku. Pokud by probíhala selekce, nebylo by možné odhadnout frekvence genotypů podle zákona Hardy-Weinberga.

Informační obsah polymorfismu (Polymorphism Informational Content - PIC) měří užitečnost sledovaného markeru (viz Rovnice 3.2). Markery s větším počtem alel či s alelami o stejné frekvenci dosahují vyšší hodnoty PIC [93].

$$PIC = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i^2 p_j^2, kde \quad (3.2)$$

i - i -tá alela

n - celkový počet alel

p_i - frekvence i -té alely

Pravděpodobnost shody (Probability of a match - PM) je také důležitým parametrem pro marker používaný ve forenzní genetice a lze ji pro jednotlivé markery spočítat dle rovnice 3.3. Tato rovnice byla sestavena Fisherem na základě nasbíraných hodnot během analýzy krevního systému MNS [94].

$$PM = \sum_{i=1}^n G_i^2, kde \quad (3.3)$$

G_i - frekvence i -tého genotypu ve zkoumaném populačním vzorku

PM pro celou sadu markerů neboli k lokusů (Combined Probability of a Match - CPM) lze vypočítat dle rovnice 3.4

$$CPM = 1 - \prod_{m=1}^k PM_m, kde \quad (3.4)$$

PM_m - pravděpodobnost shody m -tého lokusu

Dalšími parametry, které se stanovují pro identifikační markery, jsou diskriminační síla (Power of Discrimination - PD), průměrný paternitní index (Average Paternity Index - API) a průměrná pravděpodobnost vyloučení (Mean Exclusion Chance - MEC). PD lze definovat jako průměrnou pravděpodobnost, že dvě náhodně vybrané osoby budou mít jiný genotyp. V případě autozomálních lokusů je součet PM a PD roven jedné (viz Rovnice 3.5).

$$PD = 1 - PM \quad (3.5)$$

PD je v případě X-vázaných lokusů jiná pro muže (PD_M , viz Rovnice 3.6) a ženy (PD_F , viz Rovnice 3.7) [95].

$$PD_M = 1 - \sum p_i^2 \quad (3.6)$$

$$PD_F = 1 - 2(\sum p_i^2)^2 + \sum p_i^4 \quad (3.7)$$

Celková PD pro sadu k lokusů (Combined Power of Discrimination - CPD) lze vypočítat dle rovnice 3.8.

$$CPD = \prod_{m=1}^k PM_m \quad (3.8)$$

MEC je pravděpodobnost vyloučení otcovství. Výpočet MEC závisí na tom, které osoby byly testovány. Pokud se jedná o standardní trio: matka - dítě (v případě určování otcovství pomocí X-STR se jedná vždy o dceru) - nařčený otec, lze vypočítat MEC dle rovnice 3.9 [95].

$$MEC_T = 1 - \sum p_i^2 + \sum p_i^4 - (\sum p_i^2)^2 \quad (3.9)$$

Pokud je v případě standardního tria místo nařčeného otce jeho matka, lze MEC vypočítat dle rovnice Krügera [96]. Pokud matka není testována, lze vypočítat MEC pro duo (dcera-otec) pomocí rovnice 3.10 [95].

$$MEC_D = 1 - 2(\sum p_i^2)^2 + \sum p_i^3 \quad (3.10)$$

Dalším parametrem, který je důležitý pro forenzní markery je hodnota paternitního indexu neboli věrohodnostního poměru. Věrohodnostní poměr v tomto případě porovnává pravděpodobnost genotypů za předpokladu, že nařčený otec je biologickým otcem dítěte s pravděpodobností genotypů za předpokladu, že neznámý nepříbuzný muž je biologickým otcem dítěte. Věrohodnostní poměr je součástí Bayesovy věty, která je uvedena ve tvaru podílů pravděpodobností na obrázku 3.5. Bayesova věta v takovém tvaru je často využívána ve forenzní genetice v případě jak identifikace, tak i určování příbuznosti. Použití Bayesovy věty je lidsky přirozené, protože tato věta ukazuje jak lze zjistit výslednou pravděpodobnost jevu pomocí aktualizace výchozí pravděpodobnosti. Výchozí podíl pravděpodobností je v tomto případě tvořen poměrem pravděpodobností hypotéz před genetickým vyšetřením. Tato část Bayesovy věty je nejsubjektivnější a její determinace náleží soudci, jak v případech forenzní identifikace, tak i v případech určování příbuznosti. Pro forenzního vědce je nejpodstatnější část Bayesovy věty věrohodnostní poměr (Likelihood Ratio - LR), což je podíl pravděpodobností důkazů za předpokladu hypotézy obhajoby anebo obžaloby. Hodnota LR nezávisí na hodnotě výchozího podílů pravděpodobností. Na rozdíl od výchozího podílu pravděpodobností, výsledný podíl pravděpodobností je

podmíněn výsledkem genetického vyšetření. Termin „paternitní index“ nahrazuje věrohodnostní poměr v případě určování otcovství.

$$\frac{\Pr(H_p | E)}{\Pr(H_d | E)} = \frac{\Pr(H_p)}{\Pr(H_d)} \times \frac{\Pr(E | H_p)}{\Pr(E | H_d)}$$

\uparrow
Výsledný
podíl
pravděpodobností

\uparrow
Výchozí
podíl
pravděpodobností

\uparrow
Věrohodnostní
poměr (LR)

Obrázek 3.5: Bayesova věta ve tvaru podílů pravděpodobností

H_p - hypotéza obžaloby, H_d - hypotéza obhajoby, E - důkaz

V případě paternitního indexu znamená symbol „E“ výsledky genetického vyšetření, H_p - hypotéza, že nařčený otec O je biologickým otcem dítěte D z matky M, H_d - hypotéza, že neznámý nepříbuzný s nařčeným otcem muž je biologickým otcem dítěte D z matky M. Potom lze paternitní index upravit dle konkrétního případu (viz Rovnice 3.11). V této rovnici je uveden případ určování otcovství, pokud dítě má genotyp A_iA_j , matka A_iA_k a nařčený otec A_jA_l .

$$\frac{\Pr(G_D, G_M, G_O | H_p)}{\Pr(G_D, G_M, G_O | H_d)} = \frac{\Pr(G_D | G_M, G_O, H_p)}{\Pr(G_D | G_M, H_d)} = \frac{1/2 \times 1/2}{1/2 \times p_j} = \frac{1}{2 \times p_j}, kde \quad (3.11)$$

G_D - genotyp dítěte

G_M - genotyp matky

G_O - genotyp nařčeného otce

Paternitní index pro standardní tria, tzn matka dítě nařčený otec, lze vypočítat dle vzorců uvedených v práci Brennera [97] (viz Tabulka 3.9). API (Average Paternity Index) představuje průměr paternitního indexu dosaženého za využití testované sady markerů na případech s prokázaným otcovstvím. API byl vypočítán pro sadu Investigator DIPplex.

U bialelických INDEL markerů byl vypočítán počet "power-effective", neboli účinných markerů. Počet účinných markerů se rovná dvojnásobku zkoumaných markerů vynásobených průměrnou genovou diverzitou [98].

Tabulka 3.9: Výpočet paternitního indexu pro standardní tria

Dítě	Matka	Nařčený otec	Vzorec LR
q	pq	q	1/q
q	p	q	-
pq	p nebo pr	q	1/q
q	q	q	1/q
pq	p nebo pr	qr	1/2q
q	p	qr	-
q	pq	qr	1/2q
q	q	qr	1/2q
pq	pq	pq	1/(p+q)
pq	pq	q	1/(p+q)
pq	pq	qr	1/(2p+2q)
q	pq	r	0
q	q	r	0
q	p	r	-

Zdroj: [97]

3.5.4 Mezipopulační srovnání

Srovnání frekvencí markerů typů INDEL obsažených v sadě Investigator DIPplex bylo provedeno ručně pomocí χ^2 testu s použitím alelických frekvencí v německé, asijské, africké, indiánské a celoevropské populaci. Očekávané počty jedinců příslušných populací byly spočítány dle postupu zobrazeného v tabulce 3.6. Výsledný vzorec χ^2 (viz Rovnice 3.1) se skládá ze čtyř dílčích částí - alela 1 v populaci A, alela 1 v populaci B, atd.

Intrapopulační a interpopulační koeficienty genetické variability byly stanoveny s použitím F-statistiky a vzorců Wrighta [99]. F_{IS} určuje subpopulační koeficient inbreedingu a lze ho spočítat dle rovnice 3.12.

$$F_{IS} = 1 - \frac{H_{obs}}{H_{exp}}, kde \quad (3.12)$$

H_{obs} - pozorovaná heterozygotita v populaci

H_{exp} - očekávaná heterozygotita v populaci

Pomocí mezipopulačního koeficientu inbreedingu F_{ST} lze určit genetickou vzdále-

		Alela 1	Alela 2	Součet řádků (R)
Populace A	O	a	b	R1=a+b
	E	(C1×R1)/N	(C2×R1)/N	
Populace B	O	c	d	R2=c+d
	E	(C1×R2)/N	(C2×R2)/N	
Součet sloupců (C)		C1=a+c	C2=b+d	N=a+b+c+d

Obrázek 3.6: Určení očekávaných hodnot pro srovnání alelických frekvencí v různých populacích pomocí χ^2 testu

O - pozorované frekvence genotypů, *E* - frekvence genotypů odvozené pomocí zjištěných frekvencí alel dle vzorce Hardy-Weinberga pro frekvence genotypů

nost mezi sledovanými populacemi dle rovnice 3.13. Z této rovnice vyplývá, že čím více se H_S přibližuje H_T , tím více se populace podobají.

$$F_{ST} = \frac{H_T - H_S}{H_T}, kde \quad (3.13)$$

H_T - podíl heterozygotů v celkové populaci

H_S - průměrný podíl heterozygotů v subpopulacích

Ve studii sady X-STR Mentype[®] Argus X-8 byly pro srovnání mezi populacemi použity buď frekvence haplotypů (srovnání s populací Německa, Japonska a Ghany) nebo alelické frekvence (srovnání s polskou a bosenskou populací) pomocí programu Arlequin v3.5 [90]. Ve studii sady markerů X-STR Decaplex bylo mezipopulační srovnání provedeno pomocí porovnání hodnot F_{ST} pro každý marker. Srovnání české populace s populací z Ugandy, severovýchodního Španělska, Polska a Německa bylo provedeno pomocí programu Arlequin v3.5 [90].

3.5.5 Rekonstrukce haplotypů

Ve studii zabývající se výzkumem fenotypových SNP byla použita metoda rekonstrukce haplotypů v oblasti HERC2-OCA2, která, jak již bylo zmíněno, silně koreluje s pigmentací. Rekonstrukce haplotypů byla provedena za účelem zjištění, zda nějaký haplotyp asociuje s určitou barvou očí. Protože se zkoumané SNP nacházejí na au-

tozómech a soubor obsahoval vzorky od nepříbuzných osob, nebylo možné stanovit haplotypy přímo. K určení haplotypů byl použit algoritmus Excoffier-Laval-Balding (ELB) [100] v programu Arlequin v3.5 [90]. Pro provedení této analýzy bylo použito původní nastavení programu. ELB je Bayesovská metoda rekonstrukce haplotypů, která aktualizuje výchozí pravděpodobnost gametické fáze multilokusového genotypu pomocí okolních lokusů a místní LD. Pro rekonstrukci haplotypů byl trénovací soubor rozdělen do dvou skupin dle původu. Důvodem bylo to, že každá populace má jiné genotypové složení, které může mít vliv na výsledek rekonstrukce haplotypů. U každého souboru byla zvlášť provedena rekonstrukce haplotypů. Po provedení této analýzy byla data opět spojena a rozdělena dle barvy očí: světlá vs. tmavá.

3.5.6 Asociační studie

Asociační testy jsou hodně využívány v lékařské genetice pro určení genů, které mohou korelovat se zkoumaným komplexním fenotypem. Takový fenotyp může být výsledkem souhry jak faktorů genetických, tak i faktorů prostředí. Aby bylo možné pozorovat jakoukoliv asociaci, je potřeba mít minimálně dvě skupiny: ty, kteří zkoumaný fenotyp mají, a ty, kteří ho nemají. V lékařské genetice tyto skupiny nejčastěji představují pacienty a kontroly. Pro studii fenotypových markerů byly nasbírané vzorky rozděleny do dvou skupin - dle barvy očí nebo dle biogeografického původu (viz Tabulka 3.1). Výstup asociačního testu kromě vlastní p-hodnoty obsahuje poměr podílů pravděpodobností (Odds ratio - OR) [101] s mezemi spolehlivosti. Příkladem výpočtu OR pro studii fenotypových SNP je rovnice 3.14. OR se rovná podílu pravděpodobnosti výskytu alely A u světloukých osob vydělenou pravděpodobností výskytu opačné alely u světloukých osob a pravděpodobnosti výskytu alely A u tmavookých osob vydělenou pravděpodobností výskytu opačné alely u tmavookých osob. Pro porovnání se obvykle používá minoritní alela. Obor hodnot OR je $[0, \infty)$. V případě studie fenotypových SNP může $OR < 1$ svědčit o tom, že zkoumaný marker asociuje s tmavou barvou očí. $OR > 1$ naznačuje, že zkoumaný marker může asociovat se světlou barvou očí.

$$OR = \frac{\frac{Pr(A|S)}{1-Pr(A|S)}}{\frac{Pr(A|T)}{1-Pr(A|T)}}, kde \quad (3.14)$$

$Pr(A|S)$ - pravděpodobnost výskytu alely A SNP markeru u osob se světlými očima

$1-\Pr(A|S)$ - pravděpodobnost výskytu opačné alely SNP markeru u osob se světlými očima

$\Pr(A|T)$ - pravděpodobnost výskytu alely A SNP markeru u osob s tmavými očima

$1-\Pr(A|T)$ - pravděpodobnost výskytu opačné alely SNP markeru u osob s tmavými očima

Meze spolehlivosti (Confidence Intervals - CI) lze definovat jako rozsah hodnot zkoumané proměnné vytvořený ze vzorku populace tak, že do něj s jistou pravděpodobností bude patřit hodnota proměnné pro celou populaci [102]. Obvykle se volí 95% CI. Odvozování závěrů o celé populaci se provádí na základě zkoumání populačního vzorku. Správnost těchto závěrů závisí na tom, jak dobře zkoumaný vzorek reprezentuje celou populaci.

Signifikantní výsledek asociačního testu u zkoumaných markerů může mít několik způsobů interpretace. První je přímá asociace mezi markerem a fenotypem. To znamená, že zkoumaný marker způsobuje predispozici k určitému fenotypu. Druhý je nepřímá asociace. Jinými slovy, marker je v těsné blízkosti markeru, který způsobuje predispozici k určitému fenotypu. Třetí je falešně pozitivní výsledek, který může být způsoben náhodou a nebo systematickou chybou (genotypizační chyba, stratifikace populace) [103].

Asociační test byl proveden u studie fenotypových SNP markerů. Byla provedena tři porovnání. Nejprve byla zkoumána asociace vybraných SNP s barvou očí u celého trénovacího vzorku. Poté byla asociace SNP s barvou očí testována jenom u bělošských vzorků. Poslední asociační test zkoumal korelaci vybraných SNP s biogeografickým původem a byl proveden na celém trénovacím vzorku. Asociační testy byly provedeny v programu PLINK v1.07 [91].

Úskalí asociačního testu spočívá v tom, že nedokáže odhalit interakce mezi jednotlivými markery. Proto byla použita multifaktorová dimenzionální redukční analýza (Multifactor dimensionality reduction - MDR), která zkoumá korelaci kombinace zkoumaných markerů se sledovaným fenotypem.

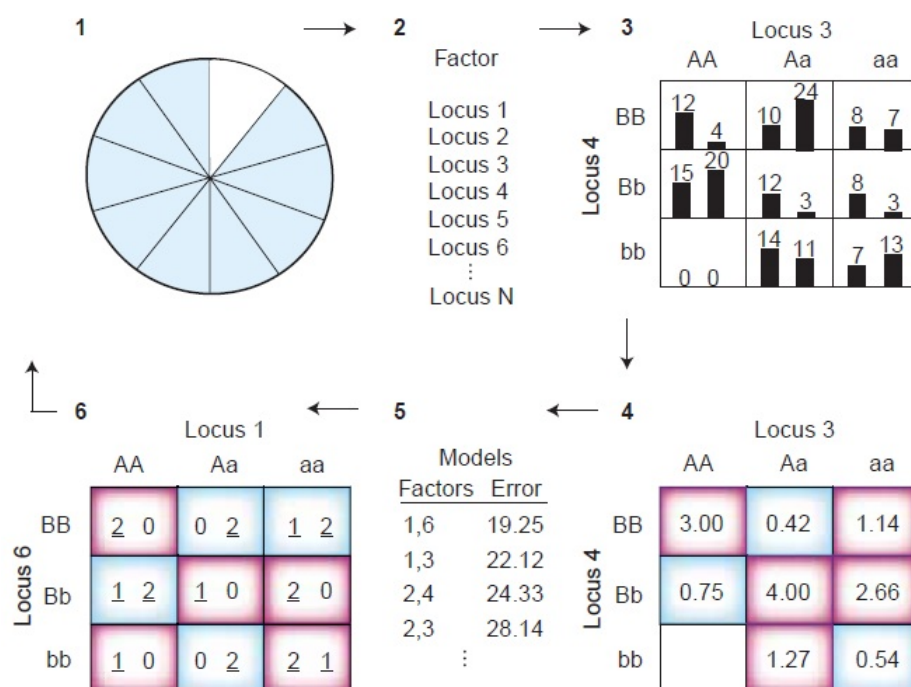
3.5.7 Multifaktorová dimenzionální redukční analýza

MDR je neparametrická metoda, které nepředpokládá určitý typ dědičnosti predispozice k projevu sledovaného fenotypu. Tato metoda je zaměřena na analýzu malého množství pozorování, u kterých je sledováno velké množství proměnných. Dle práce Ritchie et al. [104] je analýza MDR relativně odolná vůči šumu způsobenému geno-

typizační chybou a chybějícími daty. MDR se nejčastěji používá pro analýzu predispozice ke komplexnímu onemocnění, nebo obecně, ke komplexnímu fenotypu. Barva očí jistě představuje komplexní fenotyp, který je tvořen více genetickými i negenetickými faktory.

Analýzu MDR lze rozdělit na dva procesy: nejprve jsou vytvořeny modely, které jsou testovány, a následně je vybrán nejlepší model (viz Obrázek 3.7). První proces vyžaduje rozdělení dat na dvě skupiny: 9/10 dat je použito na vytvoření modelu, zbývající 1/10 na testování míry predikce tohoto modelu. Před vytvořením modelů pomocí analýzy MDR musí být nastaven rozsah počtu markerů SNP, které mohou být použity v modelu. U fenotypových SNP byly testovány kombinace od 1 do 4 markerů. Data z trénovacího souboru jsou rozdělena na „high risk“ a „low risk“ dle poměru pacientů a kontrol. Pokud tento poměr přesahuje 1, potom jsou tyto data označena jako „high risk“, pokud tento poměr nepřesahuje 1, potom jsou data označena jako „low risk“. V tomto kroku dochází k redukci dimenzí, neboť z několika proměnných, z nichž každá představuje genotyp jednotlivých markerů, se stává jedna proměnná představující multilokusový genotyp. Protože v případě barvy očí nelze ani jednu ze zkoumaných skupin označit jako pacienty nebo kontroly, byla MDR analýza provedena dvakrát: nejprve byly označeny jako pacienti světlé osoby, potom tmavé osoby. Na základě toho, jak jsou zkoumaný marker nebo kombinace markerů schopné správně rozřadit vzorky do skupiny „pacientů“ nebo „kontrol“, je určena míra klasifikační chyby. U modelu, který má nejmenší klasifikační chybu, je stanovena míra predikční chybovosti pomocí testovací skupiny dat. Výpočet klasifikační a predikční chybovosti je stejný, ale je počítán z různých datových skupin: klasifikační chybovost je počítána z trénovacího souboru, predikční chybovost z testovacího souboru. Klasifikační a predikční míra chybovosti se rovnají průměru senzitivity a specifity [105]. Proces výběru nejlepšího modelu probíhá 10 krát a nazývá se Cross Validation (CV). Dobrý model by měl mít co nejvyšší CV a nejnížší míru predikční chyby. Zatímco míra predikční chybovosti je nejmenší u nejlepšího modelu, míra klasifikační chybovosti se snižuje s počtem analyzovaných markerů. To může vést k falešně pozitivním výsledkům a k jevu zvanému „model-overfitting“. Tento jev nastává tehdy, když model popisuje šum způsobený velkou komplexností modelu místo skutečných interakcí mezi markery [106]. Aby se předešlo tomuto jevu, je před samotnou MDR analýzou potřeba použít filtr, který sníží riziko

falešně pozitivních interakcí.



Obrázek 3.7: Schéma kroků během MDR analýzy

1 - rozdělení dat do skupin: 9/10 pro určení nejlepšího modelu a 1/10 pro určení jeho míry predikce, 2 - výběr n markerů, 3 - rozdělení vzorků dle genotypů a fenotypů, 4 - determinace „high risk“ a „low risk“ genotypů, 5,6 - určení modelu s nejmenší klasifikační chybou a následné určení jeho míry predikce na nezávislých datech

Zdroj: [106]

Ve studii fenotypových SNP byl použit filtr ReliefF. Tento filtr hodnotí každý marker podle dvou parametrů: vzdálenost mezi dvěma náhodně vybranými osobami se stejným fenotypem a vzdálenost mezi dvěma náhodně vybranými osobami s opačným fenotypem. Čím menší je první vzdálenost a čím větší je druhá vzdálenost, tím je větší informativita zkoumaného markeru. Tato veličina může nabývat hodnot od -1 (nejhorší) do +1 (nejlepší) [107].

Pro určení, zda je míra predikce nejlepšího vybraného modelu statistický signifikantní, musí být následně použit permutační test. Tento test lze česky pojmenovat jako „obměňovací“. Nulová hypotéza tohoto testu je v případě studie fenotypových markerů následující: vybrané SNP markery nemají asociaci s konkrétním fenotypem.

Aby mohla tato hypotéza být zamítnuta či nezamítnuta, genotypům je náhodně přiřazen fenotyp. Potom u těchto dat probíhá analýza MDR, na jejímž základě je odhadnuta míra predikční chybovosti. Takovéto náhodné přiřazování fenotypu je opakováno 1000 krát. Potom je vytvořena distribuce míry predikční chybovosti všech 1000 pokusů. Na základě této distribuce je následně spočítána p-hodnota pro vybraný model a hladina významnosti [108]. Po provedení permutačního testu je tedy možné udělat závěr, zda vybraný model (skupina SNP markerů) správně rozřazuje vzorky do stanovených skupin se statistickou významností.

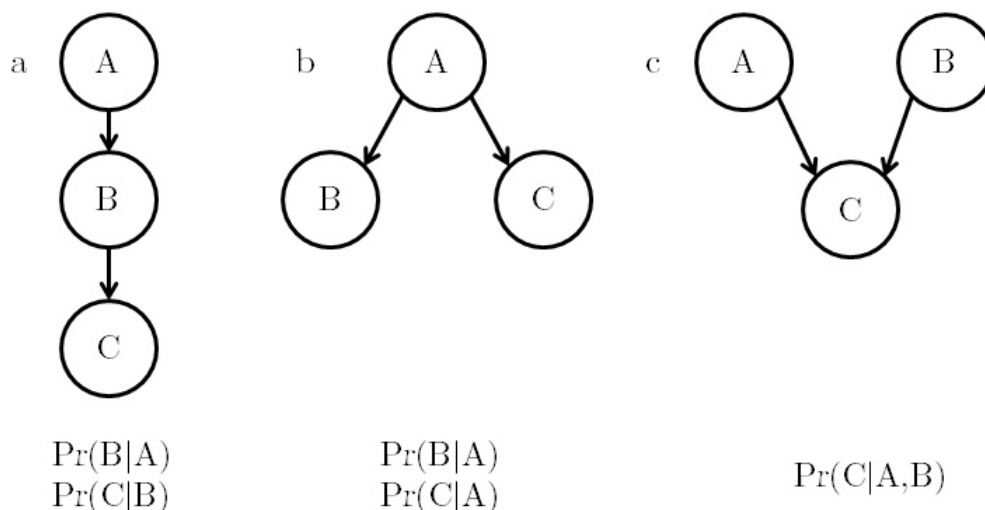
Metoda MDR byla použita ve studii fenotypových SNP na vzorky rozdělené dle barvy očí (světlá *versus* tmavá barva očí) a nebo dle původu (běloch *versus* asiat) v programu MDR (<http://www.epistasis.org>). Před provedením MDR byla data filtrována pomocí filtru ReliefF programu MDR (<http://www.epistasis.org>). Pět nejinformativnějších SNP bylo vybráno pro MDR. MDR testovalo kombinace markerů SNP obsahující od 1 do 4 markerů v modelu. Na základě této analýzy byl vybrán nejlepší model pro určování barvy očí a biogeografického původu. Oba modely byly následně testovány permutačním testem pomocí programu MDRpt package (www.epistasis.org). Pro praktickou implementaci získaných modelů byla navržena Bayesovská síť.

3.5.8 Analýza pomocí Bayesovských sítí

Bayesovské sítě (BS) představují perfektní nástroj dovolující kombinaci více zkoumaných jevů a snadnou aktualizaci odhadu jejich pravděpodobností. BS dovolují zadat výchozí pravděpodobnost nějakého jevu (v tomto případě pravděpodobnost barvy očí nebo biogeografického původu zkoumané osoby), následně zadat aktualizující informace (v tomto případě zjištěné genotypy) a v závěru vypočítat výslednou pravděpodobnost jevu se započtením aktualizujících informací. BS představují orientovaný acyklický graf, v němž šipky představují příčinná spojení. Tvorba sítí podléhá Bayesovu zákonu, který je zde uveden v nejjednodušší formě (viz rovnice 3.15). Slovně lze tuto rovnici popsat takto: pravděpodobnost A za předpokladu B se rovná pravděpodobnosti B za předpokladu A, která je vynásobena pravděpodobností A a vydělena pravděpodobností B.

$$Pr(A|B) = \frac{Pr(B|A) \times Pr(A)}{Pr(B)} \quad (3.15)$$

BS se skládá z uzlů (rodičovských a dětských) a ze šipek. Uzel je proměnná, která může nabývat různých stavů; šipky určují podmíněnost mezi proměnnými. Na obrázku 3.8 jsou zobrazena možná spojení uzlů v BS. Obrázek 3.8a představuje spojení typu „nepřímá příčina“: proměnná A je podmíněně nezávislá na proměnné C za předpokladu B. Proto, určujeme pravděpodobnost A jako $\Pr(A|B)$ a ne jako $\Pr(A|B,C)$. A lze označit jako rodičovský uzel, B jako dětský uzel ve vztahu k A a rodičovský ve vztahu k C. Druhý typ spojení je „společná příčina“ (viz Obrázek 3.8b): na proměnné A závisí jak proměnná B, tak i proměnná C. Poslední typ spojení je „společný účinek“ (viz Obrázek 3.8c): proměnná C je závislá na proměnné A a B. Proto se pravděpodobnost C určuje jako $\Pr(C|A,B)$.



Obrázek 3.8: Typy spojení uzlů v Bayesovských sítích

Proces tvorby BS zahrnuje 3 kroky. První je definování proměnných, které budou v BS. Výběr proměnných závisí na znalostech o daném tématu a úrovni detailů, na kterých má být BS vytvořena. Další krok je určování vztahů mezi proměnnými. Vztahy jsou určovány šipkami a uzly lze spojit třemi výše uvedenými způsoby. Posledním krokem je určení pravděpodobností stavů, kterých proměnné mohou nabýt. Protože pro BS platí pravidlo podmíněné nezávislosti, pravděpodobnost dětského uzlu je podmíněna jenom nejbližšími rodičovskými uzly. Takto vytvořená síť má výchozí hodnoty pravděpodobností a je připravena pro aktualizaci a následném odhadu výsledné pravděpodobnosti. V našem případě, je po vnesení důkazu v podobě výsledku genotypizace konkrétního vzorku aktualizována výchozí pravděpodobnost

barvy očí a biogeografického původu. Pomocí BS lze spočítat i věrohodnostní poměr, který určuje pravděpodobnost získaných genotypů za předpokladu dvou, vzájemně se vylučujících hypotéz (světlé vs. tmavé očí nebo běloch vs. asiát).

BS pro studii fenotypových SNP byly vytvořeny v programu Hugin Lite v7.6 (Hugin Expert A/S, Aalborg, Dánsko). BS byla vytvořena po výběru nejlepšího modelu pro určování barvy očí a biogeografického původu a byla složena z následujících uzlů: barva očí, biogeografický původ a markery SNP, které byly vybrány v rámci analýzy MDR.

3.5.9 Validace vytvořeného modelu

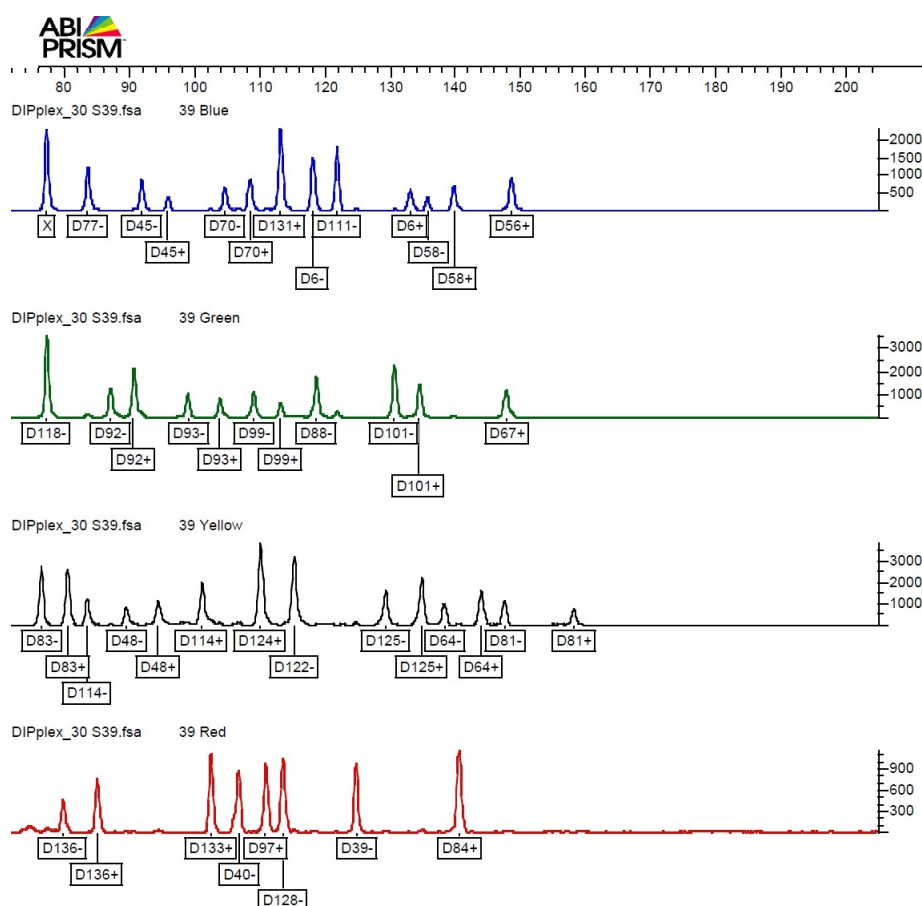
Validační studie byla provedena na 128 vzorcích: 47 Čechů, 39 Japonců žijících v Tokiu (JPT) a 42 Han Číňanů žijících v Pekingu (CHB). Pro validaci vytvořených modelů byla použita vytvořená BS. Pro určení fenotypu byla stanovena hranice výsledné pravděpodobnosti 0,7, která se ukázala jako kompromis mezi mírou chybného určení fenotypu a podílem vzorků, u nichž nebylo možné udělat závěr o fenotypu [67]. Po provedení určení fenotypu byly určeny parametry, které hodnotí robustnost BS. Pro světlou barvu očí byla stanovena senzitivita, která měří podíl skutečně světlookých osob, u kterých byla predikována světlá barva očí. Dále byla odhadnuta specifita, která určuje procento tmavookých osob, u kterých byla predikována tmavá barva očí. Pozitivní predikční hodnota (PPH) udává podíl skutečně světlookých osob ze všech osob, u kterých byla predikována světlá barva očí. Negativní predikční hodnota (NPH) je podíl skutečně tmavookých osob ze všech osob, které byly predikovány jako tmavooké. Posledním stanoveným parametrem byla přesnost modelu, která se rovná podílu osob se správně určenou barvou očí ze všech testovaných vzorků. Všechny výše uvedené parametry byly také stanoveny pro určování biogeografického původu.

4 VÝSLEDKY

4.1 Investigator DIPplex

4.1.1 Optimalizace amplifikace a fragmentační analýzy

Bylo zjištěno, že optimální množství DNA v reakci o objemu 12,5 μ l je 0,5 ng (viz Obrázek 4.1). Vyšší množství způsobuje vytažené a abnormálně tvarované píky. Bylo pozorováno selhání amplifikace lokusů nebo jednotlivých alel. Nejčastěji k vypadnutí alel a nebo lokusů docházelo u markerů rs161090, rs2307956 a rs1610937. Nebyla nalezena korelace mezi velikostí ampliconu vypadlého lokusu a frekvencí vypadnutí tohoto lokusu. Nejlepší výsledky fragmentační analýzy byly dosaženy po desetinásobném naředění směsi PCR po amplifikaci před nasazením vzorku na fragmentační analýzu.



Obrázek 4.1: Genetický profil vytvořený pomocí sady Investigator DIPplex na základě analýzy 0.5 ng templátové DNA

4.1.2 Populační parametry

Za výše uvedených podmínek byly získány genotypy pro sadu Investigator DIPplex 55 osob z České republiky. Po provedení testu na rovnováhu HWE a použití Bonferroniho korekce bylo zjištěno, že všechny zkoumané markery byly v HWE (viz Tabulka 4.1), p -hodnota byla $\geq 0,042$. Průměrná heterozygotita zkoumaného souboru byla 0,457, jejíž hodnota byla jenom nepatrně menší než genová diverzita (0,496). Intrapopulační koeficient inbreedingu byl v průměru 0,057. Provedený test LD nezjistil statisticky signifikantní vazebnou nerovnováhu mezi zkoumanými lokusy ($\alpha=0,0017$ po provedení Bonferroniho korekce).

4.1.3 Forenzní parametry

CPM pro sadu Investigator DIPplex byla 1 z $6,8 \times 10^{12}$. Tato hodnota CPM je menší než u běžně používaných sad STR markerů, ale je srovnatelná s hodnotami CPM pro ostatní INDEL sady (viz Tabulka 4.2).

Hodnoty PM, PD, PIC a API pro jednotlivé lokusy jsou představeny v tabulce 4.3. Průměrná PD byla lehce nad 60%. Tato hodnota znamená, že existuje 60% pravděpodobnost náhodného výběru dvou osob s různými genotypy téhož polymorfismu INDEL. CPD celé sady markerů dosáhla hodnoty 99,999999999%. Informační obsah (PIC) jednotlivých polymorfizmů obsažených v této sadě markerů INDEL byl v průměru 0,485, což odpovídá PIC pro bialelické markery. API byl průměrně 1,41. Kombinovaný průměrný paternitní index byl u vyšetřovaných trií přes 25000. Pravděpodobnost otcovství dosáhla 99,99%, pokud byla zvolena 50% výchozí pravděpodobnost. Ovšem pokud byla sada Investigator DIPplex použita společně se sadou markerů STR, například PowerPlex 16, dokázala zvýšit celkový paternitní index o několik řádů (viz Tabulka 4.4).

Počet účinných markerů zkoumané sady INDEL byl na základě charakteristik jednotlivých polymorfních markerů stanoven na 29,76, což se shoduje se skutečným počtem markerů obsažených v Investigator DIPlex.

Tabulka 4.1: Populační parametry markerů Investigator DIPplex v české populaci

rs číslo	p(+)	p(-)	p-hodnota*	F_{IS}	Hex	He
rs1611048	0,5455	0,4545	0,198	0,047	0,459	0,473
rs2307959	0,5455	0,4545	0,198	0,047	0,459	0,473
rs1611001	0,6182	0,3818	0,225	-0,001	0,385	0,473
rs2307652	0,4545	0,5455	0,198	0,047	0,550	0,473
rs1610905	0,4818	0,5182	0,042	0,235	0,523	0,382
rs1305047	0,5182	0,4818	0,042	0,235	0,486	0,382
rs1610937	0,6182	0,3818	0,185	0,076	0,385	0,436
rs2308292	0,6727	0,3273	0,238	0,009	0,330	0,436
rs16438	0,5273	0,4727	0,213	-0,021	0,477	0,509
rs17174476	0,4273	0,5727	0,183	0,071	0,578	0,455
rs2307570	0,5545	0,4455	0,106	0,154	0,450	0,418
rs2308163	0,6000	0,4000	0,074	-0,212	0,404	0,582
rs8190570	0,4273	0,5727	0,114	0,146	0,578	0,418
rs2307433	0,5455	0,4545	0,213	-0,027	0,459	0,509
rs1305056	0,5091	0,4909	0,192	0,054	0,495	0,473
rs2308072	0,4364	0,5636	0,077	0,187	0,569	0,400
rs2307581	0,3364	0,6636	0,238	-0,018	0,670	0,455
rs28369942	0,5273	0,4727	0,194	0,052	0,477	0,473
rs6481	0,7000	0,3000	0,225	-0,082	0,303	0,455
rs8178524	0,5091	0,4909	0,130	0,127	0,495	0,436
rs16388	0,5545	0,4455	0,174	0,080	0,450	0,455
rs1610935	0,4727	0,5273	0,132	0,125	0,532	0,436
rs17879936	0,4545	0,5455	0,071	0,193	0,550	0,400
rs16363	0,4545	0,5455	0,071	0,193	0,550	0,400
rs2067235	0,5636	0,4364	0,213	-0,035	0,440	0,509
rs17238892	0,5182	0,4818	0,210	0,017	0,486	0,491
rs2307956	0,4091	0,5909	0,193	0,060	0,596	0,455
rs2307924	0,4455	0,5545	0,197	-0,067	0,560	0,527
rs17878444	0,3182	0,6818	0,153	0,120	0,688	0,382
rs3081400	0,5091	0,4909	0,178	-0,091	0,495	0,545

* p -hodnota Fisherova testu, F_{IS} - koeficient inbreedingu, Hex - genová diverzita, He - pozorovaná heterozygotita

Tabulka 4.2: Srovnání celkové pravděpodobnosti shody různých sad markerů

Sada markerů	Typ markerů	Počet markerů	Populace	CPM
PowerPlex 16	STR	15	bělošská	1 z $1,8 \times 10^{17}$
Pimenta, Pena [46]	INDEL	40	bělošská	1 z $2,9 \times 10^{16}$
Pereira et al. [47]	INDEL	38	bělošská	1 z $2,8 \times 10^{14}$
Li et al. [48]	INDEL	29	asijská	1 z $1,1 \times 10^{11}$
Tato studie	INDEL	30	bělošská (Češi)	1 z $6,8 \times 10^{12}$

CPM - celková pravděpodobnost shody

Tabulka 4.3: Forenzní parametry markerů DIPplex v české populaci

rs číslo	PM	PD	PIC	API
rs1611048	0,367	0,633	0,496	1,41
rs2307959	0,367	0,633	0,496	1,44
rs1611001	0,390	0,610	0,472	1,41
rs2307652	0,367	0,633	0,496	1,24
rs1610905	0,338	0,662	0,499	1,54
rs1305047	0,338	0,662	0,499	1,56
rs1610937	0,377	0,623	0,472	1,34
rs2308292	0,409	0,591	0,440	1,13
rs16438	0,381	0,619	0,499	1,27
rs17174476	0,366	0,634	0,489	1,59
rs2307570	0,350	0,650	0,494	1,32
rs2308163	0,446	0,554	0,480	1,25
rs8190570	0,355	0,645	0,489	1,23
rs2307433	0,384	0,616	0,496	1,41
rs1305056	0,363	0,637	0,500	1,45
rs2308072	0,348	0,652	0,492	1,45
rs2307581	0,409	0,591	0,446	1,27
rs28369942	0,364	0,636	0,499	1,61
rs6481	0,435	0,565	0,420	1,71
rs8178524	0,349	0,651	0,500	1,28
rs16388	0,361	0,639	0,494	1,77
rs1610935	0,351	0,649	0,499	1,56
rs17879936	0,344	0,656	0,496	1,53
rs16363	0,344	0,656	0,496	1,66
rs2067235	0,388	0,612	0,492	1,27
rs17238892	0,371	0,629	0,499	1,34
rs2307956	0,372	0,628	0,483	1,51
rs2307924	0,396	0,604	0,494	1,20
rs17878444	0,403	0,597	0,434	1,51
rs3081400	0,401	0,599	0,500	1,18

PM - pravděpodobnost náhodné shody, PD - diskriminační síla, PIC - informační obsah polymorfizmu, API - průměrný paternitní index

Tabulka 4.4: Příklad použití kombinací markerů INDEL a STR pro určení otcovství

Sada markerů	CPI	W (%)
PowerPlex16	54116364	99,99999
DIPplex	49366	99,99
Combined	$2,67 \times 10^{12}$	99,9999999999

CPI - celkový paternitní index, W - pravděpodobnost otcovství za předpokladu 50% výchozí pravděpodobnosti

4.1.4 Mezipopulační porovnání

Výsledky studie české populace byly srovnány s německou, indiánskou, japonskou a africkou populací. Alelické frekvence markerů ze sady Investigator DIPplex byly pro tyto populace získány ze stránek výrobce (<http://www.qiagen.com>). V tabulce 4.5 jsou hodnoty χ^2 testu (hladina významnosti po Bonferroniho korekci - 0,002). Z výsledků je patrné, že se alelické frekvence statisticky významně neliší u české a německé populace. Největší rozdíly ve frekvenci alel jsou mezi českou a indiánskou populací. Celkem zde bylo nalezeno 5 lokusů, jejichž frekvence se statisticky významně lišila u osob z české a indiánské populace: rs2307959, rs1610937, rs2308072, rs16388 a rs1630935. Porovnání české a africké populace vedlo ke zjištění 4 lokusů s odlišnými frekvencemi: rs1610937, rs2307581, rs6481 a rs1610935. Mezi českou a japonskou populací byl nalezen statisticky významný rozdíl ve frekvencích alel ve dvou lokusech: rs1305047 a rs2307956.

Tabulka 4.5: Mezipopulační srovnání alelických frekvencí markerů obsažených v sadě Investigator DIPplex - hodnoty χ^2 testu

číslo rs	CZ-GE	CZ-AMI	CZ-JA	CZ-AF
rs1611048	0,0381	0,0448	2,6290	7,7671
rs2307959	0,0733	12,7591	1,0822	3,6698
rs1611001	0,7336	5,7277	1,7534	0,1756
rs2307652	0,5417	6,4793	4,1583	9,6639
rs1610905	0,6073	0,0698	1,2117	0,0886
rs1305047	0,0762	0,1207	10,8199	0,5862
rs1610937	1,4880	12,6169	0,3321	19,0390
rs2308292	0,0408	0,9534	0,3996	0,0005
rs16438	1,0417	8,0162	10,4117	7,9211
rs17174476	0,0683	-	-	-
rs2307570	0,0027	4,0793	0,7887	2,9769
rs2308163	0,8647	9,9415	5,8006	4,2899
rs8190570	0,0371	-	-	-
rs2307433	0,0003	-	-	-
rs1305056	5,0535	3,0442	0,0001	2,0177
rs2308072	0,2416	11,8779	0,0009	1,8346
rs2307581	0,1515	2,4281	0,3035	16,0679
rs28369942	0,0678	-	0,9621	0,0737
rs6481	0,8119	8,2371	0,1280	28,9520
rs8178524	1,7514	-	-	2,0519
rs16388	0,0003	19,0629	1,4369	9,5765
rs1610935	0,5388	11,8788	8,5160	12,0451
rs17879936	0,2033	-	-	-
rs16363	0,2033	0,3922	2,9061	6,0916
rs2067235	0,0508	0,9697	8,1810	0,3650
rs17238892	0,0050	-	-	0,6226
rs2307956	0,0132	1,6532	11,6738	0,7685
rs2307924	0,3081	2,4906	0,6448	2,5434
rs17878444	0,6899	-	-	-
rs3081400	0,9534	-	-	-

statisticky významné hodnoty χ^2 testu, CZ - česká populace, GE - německá populace, AM - američtí indiáni, JA - japonská populace, AF - africká populace

4.2 Mentype[®] Argus X-8

4.2.1 Populační parametry

Test HWE, který byl proveden na souboru obsahujícím ženské vzorky, neukázal žádné statisticky významné odchylky (viz Tabulka 4.6). Všechny hodnoty byly vyšší než hladina významnosti, která byla 0,00625 po Bonferroniho korekci. Pozorována heterozygozita byla v průměru 0,8131. LD neodhalil statisticky významnou vazebnou nerovnováhu mezi zkoumanými markery. Hodnota významnosti byla pro tento test stanovena na 0,00178 po Bonferroniho korekci.

Tabulka 4.6: Výsledky testu HWE pro markery ze sady Argus X-8

	Poz. Het.	Oč. Het	p-hodnota
DXS10135	0,9596	0,94011	0,80791
DXS8378	0,67677	0,69656	0,24928
DXS7132	0,75758	0,74348	0,49808
DXS10074	0,80808	0,829	0,08214
HPRTB	0,78788	0,7605	0,73073
DXS10101	0,89899	0,89781	0,34727
DXS10134	0,84848	0,85992	0,66456
DXS7423	0,76768	0,70164	0,36212

Poz. Het. (O) - pozorovaná heterozygozita, Oč. Het. (E)- očekávaná heterozygozita

V tabulce 4.7 jsou alelické frekvence markerů ze sady Argus X-8 u souboru obsahujícího 99 žen a 78 mužů. V lokusu DXS10135 bylo zjištěno nejvíce alel -25, DXS7423 a DXS8378 měly nejméně - 5. Pomocí této studie byly poprvé nalezeny následující alely: alela 15.3 v DXS10135, 30.2 a 37.1 v DXS10134.

Frekvence haplotypů jsou uvedeny v tabulce 4.8. Tyto frekvence byly spočítány ze vzorků 78 mužů, protože u nichž haplotyp odpovídá jejich jedinému chromozómu X. Nejvíce zjištěných haplotypů (37) bylo v první vazebné skupině (Linkage group - LG1). Nejčastější haplotypy v LG1 byly 25-12 a 24-12, v LG2 13-16, v LG3 13-31.2 a v LG4 35-15.

Tabulka 4.7: Alelické frekvence markerů ze sady Argus X-8

Alela	DXS10135	DXS8378	DXS7132	DXS10074	HPRTB	DXS10101	DXS10134	DXS7423
7	-	-	-	0,03986	0,02536	-	-	-
8	-	-	-	0,15217	-	-	-	-
9	-	0,01449	-	0,00725	-	-	-	-
10	-	0,32971	-	-	0,01087	-	-	-
11	-	0,36957	0,00362	0,00362	0,1087	-	-	-
11.2	-	-	-	-	0,00362	-	-	-
12	-	0,24638	0,1087	0,00362	0,31884	-	-	-
13	-	0,03986	0,32609	0,01087	0,30797	-	-	0,09783
14	-	-	0,33333	0,01087	0,19203	-	-	0,31884
15	-	-	0,17391	0,06884	0,02536	-	-	0,41667
15.3	0,00362	-	-	-	-	-	-	-
16	0,00362	-	0,04348	0,22464	0,00725	-	-	0,13406
17	0,01087	-	0,01087	0,24638	-	-	-	0,03261
18	0,04348	-	-	0,16667	-	-	-	-
19	0,06159	-	-	0,05797	-	-	-	-
19.1	0,00362	-	-	-	-	-	-	-
20	0,04348	-	-	0,00725	-	-	-	-
20.1	0,03261	-	-	-	-	-	-	-
21	0,06159	-	-	-	-	-	-	-
21.1	0,02174	-	-	-	-	-	-	-
22	0,05797	-	-	-	-	-	-	-
22.1	0,00725	-	-	-	-	-	-	-
23	0,08333	-	-	-	-	-	-	-
23.1	0,01449	-	-	-	-	-	-	-
24	0,08696	-	-	-	-	-	-	-
25	0,0942	-	-	-	-	-	-	-
25.1	0,00725	-	-	-	-	-	-	-
25.2	-	-	-	-	-	0,01449	-	-
26	0,0942	-	-	-	-	-	-	-
26.2	-	-	-	-	-	0,00362	-	-
27	0,09783	-	-	-	-	0,00362	-	-
27.2	-	-	-	-	-	0,05072	-	-
28	0,05072	-	-	-	-	0,02174	-	-
28.2	-	-	-	-	-	0,12319	-	-
29	0,0471	-	-	-	-	0,06884	-	-
29.2	-	-	-	-	-	0,12319	-	-
30	0,03986	-	-	-	-	0,02899	0,00362	-
30.2	-	-	-	-	-	0,16304	0,00362	-
31	0,02174	-	-	-	-	0,08333	0,01087	-
31.2	-	-	-	-	-	0,11232	-	-
32	0,00725	-	-	-	-	0,07609	0,00362	-
32.2	-	-	-	-	-	0,0471	-	-
33	-	-	-	-	-	0,0471	0,05435	-
33.2	-	-	-	-	-	0,01087	-	-
34	0,00362	-	-	-	-	0,01812	0,12681	-
34.2	-	-	-	-	-	0,00362	-	-
35	-	-	-	-	-	-	0,19928	-
35.2	-	-	-	-	-	-	0,00362	-
36	-	-	-	-	-	-	0,20652	-
37	-	-	-	-	-	-	0,17029	-
37.1	-	-	-	-	-	-	0,00362	-
37.2	-	-	-	-	-	-	0,00725	-
37.3	-	-	-	-	-	-	0,00725	-
38	-	-	-	-	-	-	0,06159	-
38.3	-	-	-	-	-	-	0,02536	-
39	-	-	-	-	-	-	0,01087	-
39.3	-	-	-	-	-	-	0,03986	-
40	-	-	-	-	-	-	0,00362	-
40.3	-	-	-	-	-	-	0,02174	-
41	-	-	-	-	-	-	0,00362	-
41.3	-	-	-	-	-	-	0,01449	-
42.3	-	-	-	-	-	-	0,01087	-
43.3	-	-	-	-	-	-	0,00725	-

Tabulka 4.8: Frekvence haplotypů markerů ze sady Argus X-8

LG1		Frek.	LG2		Frek.	LG3		Frek.	LG4		Frek.
17	12	0,0128	11	16	0,0128	9	30,2	0,0256	31	15	0,0128
18	10	0,0256	12	15	0,0128	10	30,2	0,0128	32	17	0,0128
18	12	0,0256	12	17	0,0256	11	28,2	0,0128	33	15	0,0385
19	10	0,0256	12	18	0,0513	11	29,2	0,0513	33	16	0,0128
19	11	0,0128	12	19	0,0128	11	30	0,0256	33	17	0,0128
19	12	0,0385	13	8	0,0513	11	30,2	0,0128	34	14	0,0385
20.1	10	0,0385	13	15	0,0256	11	31,2	0,0256	34	15	0,0385
20.1	12	0,0128	13	16	0,1154	11	32,2	0,0128	34	16	0,0769
21	10	0,0513	13	17	0,0641	12	25,2	0,0128	34	17	0,0128
21	12	0,0128	13	18	0,0385	12	27,2	0,0256	35	14	0,0641
21.1	10	0,0256	13	19	0,0256	12	28,2	0,0256	35	15	0,1026
22	10	0,0128	14	7	0,0513	12	29	0,0128	35	16	0,0385
22	11	0,0128	14	8	0,0513	12	29,2	0,0641	36	14	0,0128
22	12	0,0128	14	13	0,0256	12	30,2	0,0513	36	13	0,0769
23	10	0,0256	14	15	0,0385	12	31	0,0256	36	15	0,0769
23	11	0,0256	14	16	0,0513	12	32	0,0128	37	13	0,0128
23	12	0,0256	14	17	0,0769	12	32,2	0,0385	37	14	0,0641
23.1	10	0,0128	14	18	0,0385	12	34	0,0128	37	15	0,0513
23.1	11	0,0128	14	19	0,0256	13	28,2	0,0513	37.1	15	0,0128
24	10	0,0128	14	20	0,0128	13	29,2	0,0641	37.3	15	0,0128
24	11	0,0385	15	8	0,0128	13	30,2	0,0128	38	13	0,0128
24	12	0,0769	15	14	0,0128	13	31	0,0128	38	14	0,0256
25	10	0,0513	15	16	0,0128	13	31,2	0,1026	38	15	0,0256
25	11	0,0128	15	17	0,0513	13	32	0,0513	38.3	14	0,0128
25	12	0,0769	15	18	0,0256	13	32,2	0,0256	38.3	16	0,0256
26	11	0,0641	15	19	0,0128	13	33	0,0641	39	13	0,0128
26	12	0,0128	16	8	0,0128	14	28	0,0128	39.3	14	0,0128
27	10	0,0256	16	17	0,0256	14	29	0,0128	39.3	15	0,0641
27	11	0,0385	17	8	0,0128	14	29,2	0,0385	40.3	16	0,0128
27	12	0,0128	17	17	0,0128	14	30	0,0128	42.3	17	0,0128
28	10	0,0256	-	-	-	14	30,2	0,0256	-	-	-
28	11	0,0128	-	-	-	14	31	0,0128	-	-	-
28	12	0,0128	-	-	-	14	31,2	0,0128	-	-	-
28	13	0,0128	-	-	-	14	32	0,0128	-	-	-
29	11	0,0256	-	-	-	15	31	0,0128	-	-	-
30	11	0,0513	-	-	-	-	-	-	-	-	-
31	11	0,0128	-	-	-	-	-	-	-	-	-

LG1 - první vazebná skupina DXS10135 - DXS8378, LG2 - druhá vazebná skupina DXS7132 - DXS10074, LG3 - třetí vazebná skupina HPRTB - DXS10101, LG4 - čtvrtá vazebná skupina DXS10134 - DXS7423, Frek. - frekvence haplotypů

4.2.2 Forenzní parametry

Hodnoty forenzních parametrů markerů ze sady Argus X-8 zjištěných v populaci středního Chorvatska jsou uvedeny v tabulce 4.9. Marker DXS10135 měl nejvyšší hodnotu PIC - 0,9306, zatímco nejnižší PIC měly markery s nejmenším počtem alel - DXS7423 (PIC=0.6447) a DXS8378 (PIC=0,6316). PIC se pro celou sadu markerů dosahoval hodnoty 0,999998. PD dosahovala u mužů hodnot od 0,6922 do 0,9345, u žen od 0,8447 do 0,9918. CPD celého souboru markerů dosáhla hodnoty 99,9999% u mužských vzorků a 99,99999999% u ženských vzorků.

Tabulka 4.9: Forenzní parametry markerů ze sady Argus X-8

Marker	PD_M	PD_F	PIC	MEC_{DEF}	MEC_T	MEC_D
DXS10135	0,9345	0,9918	0,9306	0,8670	0,9304	0,8741
DXS8378	0,6922	0,8447	0,6316	0,4230	0,6316	0,4857
DXS7132	0,7385	0,8879	0,6948	0,5055	0,6948	0,5553
DXS10074	0,8278	0,9483	0,8058	0,6596	0,8058	0,6907
HPRTB	0,7533	0,8992	0,7133	0,5295	0,7133	0,5767
DXS10101	0,9040	0,983 0	0,8962	0,8061	0,8960	0,8192
DXS10134	0,8623	0,9666	0,8479	0,7271	0,8476	0,7488
DXS7423	0,6961	0,8562	0,6447	0,4483	0,6447	0,5004

PD_M - diskriminační síla pro muže, PD_F - diskriminační síla pro ženy, PIC - informační obsah polymorfizmu, MEC_{DEF} - průměrná pravděpodobnost vyloučení pro deficientní tria, MEC_T - průměrná pravděpodobnost vyloučení pro tria, MEC_D - průměrná pravděpodobnost vyloučení pro dua

4.2.3 Mezipopulační porovnání

Haplotypové frekvence středochorvatské populace byly pro každou vazebnou skupinu srovnány s haplotypovými frekvencemi populace Ugandy, Japonska a Německa [109] (viz Tabulka 4.10). Hladina významnosti byla 0,0125 po Bonferroniho korekci. Statisticky významné odchylky ve frekvencích haplotypů byly nalezeny u všech LG po porovnání středochorvatské a japonské populace. Porovnání středochorvatské populace a populace Ghany odhalilo statisticky významnou genetickou vzdálenost

u LG2 a LG3. Po porovnání mezi středochovatskou a německou populací nebyly nalezeny žádné statisticky významné odchylky ve frekvencích haplotypů.

Tabulka 4.10: Srovnání haplotypových frekvencí markerů ze sady Argus X-8

LG	Japonsko [109]		Ghana [109]		Německo [109]	
	F_{ST}	p-hodnota	F_{ST}	p-hodnota	F_{ST}	p-hodnota
1	0,00869	0,00901*	0,00697	0,54050	0,00454	0,02703
2	0,01495	0,00000*	0,01554	0,00000*	-0,00056	0,48649
3	0,01058	0,00901*	0,01329	0,00000*	0,00471	0,01802
4	0,02273	0,00000*	0,00686	0,03604	0,00386	0,05405

* statisticky významné hodnoty, LG - vazebná skupina, F_{ST} - mezipopulační genetická vzdálenost

Bylo provedeno srovnání alelických frekvencí mezi středochovatskou, jihopolskou a bosenskou populací (viz Tabulka 4.11). Polská populace byla vybrána, neboť patří - stejně jako středochovatská - do slovanské populace. Bosna a Hercegovina sdílí s Chorvatskem společnou hranici. Test LD neodhalil žádnou statisticky významnou vazebnou nerovnováhu mezi středochovatskou, polskou a bosenskou populací ($\alpha=0,006$ po Bonferroniho korekci).

Tabulka 4.11: Srovnání alelických frekvencí markerů ze sady Argus X-8

	Polsko [110]		Bosna [111]	
	F_{ST}	p-hodnota	F_{ST}	p-hodnota
DXS10135	0.00001	0.44434	-	-
DXS8378	-0.00057	0.47754	-0.00275	0.68457
DXS7132	-0.00088	0.52637	0.00104	0.31055
DXS10074	-0.00224	0.88965	-	-
HPRTB	-0.00035	0.42090	0.0013	0.28516
DXS10101	-0.00032	0.50879	-	-
DXS10134	-0.00141	0.74219	-	-
DXS7423	-0.00211	0.76562	-0.0032	0.78223

F_{ST} - mezipopulační genetická vzdálenost

4.3 Decaplex X-STR

4.3.1 Populační parametry

V tabulce 4.12 jsou představeny alelické frekvence markerů ze sady Decaplex X-STR. U markerů bylo detekováno v průměru 7 alel. Nejmenší počet alel byl 6 (DXS8378, DXS7133, DXS7423), největší 12 (DXS6809, DXS6789).

Tabulka 4.12: Alelické frekvence markerů ze sady Decaplex X-STR

Alela	DXS8378	DXS9898	DXS7133	GATA31E08	GATA172D05	DXS7423	DXS6809	DXS7132	DXS9902	DXS6789
6	-	-	-	-	0,144	-	-	-	-	-
7	-	-	0,005	-	-	-	-	-	-	-
8	-	-	-	0,001	0,196	-	-	-	-	-
8.3	-	0,277	-	-	-	-	-	-	-	-
9	0,019	-	0,471	0,200	0,028	-	-	-	0,001	-
10	0,367	0,003	0,132	0,041	0,287	0,001	-	-	0,048	-
11	0,333	0,227	0,341	0,222	0,224	-	-	0,006	0,343	-
12	0,258	0,284	0,038	0,185	0,116	-	-	0,093	0,344	-
13	0,020	0,15	0,014	0,242	0,004	0,107	-	0,297	0,251	-
14	0,005	0,051	-	0,100	-	0,347	-	0,356	0,011	0,009
15	-	0,007	-	0,010	-	0,367	-	0,201	0,001	0,053
16	-	-	-	-	-	0,151	-	0,043	-	0,025
17	-	-	-	-	-	0,026	-	0,003	-	0,006
18	-	-	-	-	-	-	-	-	-	0,002
19	-	-	-	-	-	-	-	-	-	0,017
20	-	-	-	-	-	-	-	-	-	0,408
21	-	-	-	-	-	-	-	-	-	0,240
22	-	-	-	-	-	-	-	-	-	0,160
23	-	-	-	-	-	-	-	-	-	0,071
24	-	-	-	-	-	-	-	-	-	0,006
25	-	-	-	-	-	-	-	-	-	0,001
27	-	-	-	-	-	-	0,021	-	-	-
28	-	-	-	-	-	-	0,043	-	-	-
29	-	-	-	-	-	-	0,016	-	-	-
30	-	-	-	-	-	-	0,012	-	-	-
31	-	-	-	-	-	-	0,162	-	-	-
32	-	-	-	-	-	-	0,150	-	-	-
33	-	-	-	-	-	-	0,284	-	-	-
34	-	-	-	-	-	-	0,194	-	-	-
35	-	-	-	-	-	-	0,066	-	-	-
36	-	-	-	-	-	-	0,036	-	-	-
37	-	-	-	-	-	-	0,014	-	-	-
38	-	-	-	-	-	-	0,001	-	-	-

Po provedení testu na HWE bylo zjištěno, že ani jeden z markerů ze sady X-STR Decaplex, nevykazoval odchylku od HWE (viz Tabulka 4.13), hladina významnosti byla 0,005 po Bonferroniho korekci.

Tabulka 4.13: Výsledky testu HWE pro markery ze sady Decaplex X-STR

Marker	Poz.Het.	Oč.Het.	p-hodnota
DXS8378	0,71066	0,69711	0,03267
DXS9898	0,78173	0,76048	0,57221
DXS7133	0,61421	0,63669	0,45710
GATA31E08	0,79188	0,80441	0,62116
GATA172D05	0,77665	0,79569	0,93628
DXS7423	0,73604	0,72340	0,18808
DXS6809	0,75635	0,82365	0,55448
DXS7132	0,74619	0,72950	0,72457
DXS9902	0,72589	0,70203	0,10650
DXS6789	0,70051	0,71895	0,72168

Poz. Het. (O) - pozorovaná heterozygotita, Oč. Het. (E)- očekávaná heterozygotita

Test vazebné nerovnováhy neukázal u mužských vzorků statisticky významnou vazbu mezi žádným párem lokusů. U ženských vzorků byla nalezena vazebná nerovnováha mezi lokusy DXS6809 a DXS6789 ($p=0,0000$). Hladina významnosti u obou testů byla 0,0011 po Bonferonniho korekci. Tento výsledek mohl být způsoben skutečností, že vzdálenost mezi oběma markery je 500Kb (viz Tabulka 4.14).

4.3.2 Forenzní parametry

V tabulce 4.15 jsou představeny nejdůležitější forenzní parametry pro každý z markerů ze sady Decaplex X-STR. Průměrná PD_M byla 0,740579, průměrná PD_F - 0,885909. Největší diskriminační sílu měl marker DXS6809, jehož PD_F dosáhla hodnoty 0,947488, PD_M byla stanovena na 0,824425. Nejmenší diskriminační sílu měl marker DXS7133: $PD_F=0,808208$, $PD_M=0,643052$. Celková PD_F byla 0,999999999, celková PD_M byla 0,999998903. MEC pro tria byla v rozsahu od 0,802740 (DXS6809) do 0,578671 (DXS7133). Celková MEC pro tria byla 0,999994966. MEC_D pro dua byla v rozmezí 0,432549 – 0,687116. Celková MEC_D dosáhla hodnoty 0,999772143.

4.3.3 Mezipopulační porovnání

V tabulce 4.16 jsou výsledky srovnání české populace s jinými populacemi (populace z Ugandy [112], severovýchodního Španělska [113], Německa [114] a Polska [115]). Srovnání interpopulační vzdálenosti mezi českou populací a populací

Tabulka 4.14: Lokalizace markerů ze sady Decaplex X-STR na chromozómu X

Marker	Lokalizace (Mb)*
DXS8378	9,37
DXS9902	15,323
DXS7132	64,655
DXS9898	87,796
DXS6809	94,938
DXS6789	95,449
DXS7133	109,041
GATA172D05	113,175
GATA31E08	140,234
DXS7423	149,668

* dle GRCh37.p10 (Genome Reference Consortium Human)

Tabulka 4.15: Forenzní parametry markerů ze sady Decaplex X-STR

	MEC_T	MEC_D	PD_M	PD_F
DXS8378	0,624845	0,478593	0,687624	0,839643
DXS9898	0,726698	0,591219	0,765927	0,905981
DXS7133	0,578671	0,432549	0,643052	0,808208
GATA31E08	0,777912	0,653596	0,806628	0,933891
GATA172D05	0,763020	0,635072	0,794047	0,926556
DXS7423	0,659151	0,515848	0,709975	0,865062
DXS6809	0,802740	0,687116	0,824425	0,947488
DXS7132	0,688967	0,548543	0,734111	0,884159
DXS9902	0,639556	0,49416	0,698587	0,85012
DXS6789	0,706264	0,569194	0,741414	0,897984

MEC_T - průměrná pravděpodobnost vyloučení pro tria, MEC_D - průměrná pravděpodobnost vyloučení pro dua, PD_M - diskriminační síla pro muže, PD_F - diskriminační síla pro ženy

Ugandy odhalilo signifikantní rozdíly ($p < 0.05$) ve všech lokusech kromě DXS7423. Oproti tomu populace severovýchodního Španělska se statisticky významně lišila od české populace jenom ve čtyřech lokusech z deseti: DXS9898, DXS7133, DXS7423 a DXS9902. Následně byla česká populace srovnána s geneticky více příbuznými populacemi - německou a polskou. Srovnání české a německé populace odhalilo statis-

tický významnou genetickou vzdálenost jenom v jednom lokusu ze třech (DXS7133). Srovnání s polskou populací neodhalilo žádnou statisticky významnou genetickou vzdálenost.

Tabulka 4.16: Porovnání genetických vzdáleností mezi českou populací a ostatními populacemi

Marker	Uganda [112]		SV Španělsko [113]		Polsko [115]		Německo [114]	
Lokus	F_{ST}	p-hodnota	F_{ST}	p-hodnota	F_{ST}	p-hodnota	F_{ST}	p-hodnota
DXS8378	0,02782	0,00000	-0,00065	0,46847	-0,00093	0,54054	0,00005	0,36937
DXS9898	0,04516	0,00000	0,00741	0,02703	-	-	0,00210	0,17117
DXS7133	0,19500	0,00000	0,00763	0,03604	-	-	0,01502	0,00901
GATA31E08	0,01965	0,00000	-0,00144	0,62162	-	-	-	-
GATA172D05	0,08427	0,00000	0,00008	0,40541	-	-	-	-
DXS7423	0,00306	0,12613	0,00744	0,02703	0,00016	0,33333	-	-
DXS6809	0,02119	0,00000	0,00107	0,23423	-	-	-	-
DS7132	0,00545	0,03604	0,00132	0,27928	0,00008	0,38739	-	-
DXS9902	0,01295	0,00901	0,00731	0,03604	-	-	-	-
DXS6789	0,08795	0,00000	-0,00266	0,90090	-	-	-	-

F_{ST} - mezipopulační genetická vzdálenost, SV Španělsko - severovýchodní Španělsko

4.4 Fenotypové SNP

4.4.1 Populační parametry

Tato studie zjistila, že marker rs6119471 je monomorfní v bělošské a asijské populaci. Dále bylo zjištěno, že žádná osoba ze 100 vyšetřených bělochů neměla genotyp GG v markeru rs1426654, ovšem byly nalezeny 2 heterozygoti AG. Test HWE, který byl proveden zvlášť pro bělošské a asijské vzorky, neodhalil žádné statisticky významné odchylky (viz Tabulka 4.17). Hladina významnosti tohoto testu byla stanovena na 0,0023 po Bonferroniho korekci. Nejnižší p-hodnota byla 0,005 u rs7495174 a rs12913832 v bělošské populaci.

Tabulka 4.17: Výsledky testu HWE pro fenotypové markery

rs číslo	Populace	Poz. Het	Oč. Het	p-hodnota
rs16891982	bělošská	0,16000	0,14720	1,00000
rs16891982	asijská	0,48390	0,44800	1,00000
rs12203592	bělošská	0,14000	0,14720	0,47860
rs12203592	asijská	0,06452	0,06243	1,00000
rs35264875	bělošská	0,32000	0,2952	0,73230
rs35264875	asijská	0,19350	0,17480	1,00000
rs1393350	bělošská	0,33000	0,32590	1,00000
rs1393350	asijská	0,09677	0,09209	1,00000
rs12821256	bělošská	0,22000	0,19580	0,60050
rs12821256	asijská	0,03226	0,09209	0,04918
rs2031526	bělošská	0,28000	0,29520	0,51540
rs2031526	asijská	0,74190	0,49170	0,00968
rs12896399	bělošská	0,48000	0,46620	0,83310
rs12896399	asijská	0,38710	0,38290	1,00000
rs1800407	bělošská	0,16000	0,14720	1,00000
rs1800407	asijská	0,06452	0,12070	0,09753
rs7495174	bělošská	0,19000	0,27560	0,00462
rs7495174	asijská	0,16130	0,24820	0,09300
rs12913832	bělošská	0,33000	0,46360	0,00477
rs12913832	asijská	0,48390	0,42510	0,67800
rs916977	bělošská	0,31000	0,32590	0,55210
rs916977	asijská	0,51610	0,49950	1,00000
rs1426654	bělošská	0,02000	0,01980	1,00000
rs1426654	asijská	0,41940	0,49170	0,46920
rs4778138	bělošská	0,36000	0,32000	0,34710
rs4778138	asijská	0,51610	0,48700	1,00000
rs1545397	bělošská	0,14000	0,13020	1,00000
rs1545397	asijská	0,54840	0,49790	0,72510
rs4778241	bělošská	0,36000	0,37500	0,78890
rs4778241	asijská	0,48390	0,49170	1,00000
rs1667394	bělošská	0,31000	0,33760	0,3837
rs1667394	asijská	0,48390	0,50000	1,00000
rs258322	bělošská	0,17000	0,15560	1,00000
rs258322	asijská	0,54840	0,49170	0,71800
rs1805007	bělošská	0,09000	0,08595	1,00000
rs1805007	asijská	0,00000	0,00000	1,00000
rs885479	bělošská	0,12000	0,11280	1,00000
rs885479	asijská	0,48390	0,49790	1,00000
rs4911414	bělošská	0,43000	0,44560	0,82210
rs4911414	asijská	0,32260	0,31220	1,00000
rs1015362	bělošská	0,42000	0,42000	1,00000
rs1015362	asijská	0,35480	0,33140	1,00000
rs6119471	bělošská	0,00000	0,00000	1,00000
rs6119471	asijská	0,00000	0,00000	1,00000

Poz. Het. (O) - pozorovaná heterozygotita, Oč. Het. (E)- očekávaná heterozygotita

Tabulka 4.18: Výsledky analýzy LD pro markery umístěné na 15. chromozómu

SNP	rs916977	rs12913832	rs1545397	rs1667394	rs1800407	rs4778138	rs4778241	rs7495174
rs916977	x	+	-	+	+	+	+	-
rs12913832	+	x	-	+	-	-	-	+
rs1545397	-	-	x	+	-	-	+	-
rs1667394	+	+	+	x	+	-	+	-
rs1800407	+	-	-	+	x	-	-	-
rs4778138	+	-	-	-	-	x	+	-
rs4778241	+	-	+	+	-	+	x	-
rs7495174	-	+	-	-	-	-	-	x

+ statisticky významný výsledek ($p < 0,0002$)

Tabulka 4.19: Vzdálenost mezi markery SNP umístěné na 15. chromozómu v oblasti OCA2-HERC2

Marker	Pozice	Vzdálenost (Kb)*
rs1545397	28187772	0
rs1800407	28230318	42,546
rs4778138	28335820	105,502
rs4778241	28338713	2,893
rs7495174	28344238	5,525
rs12913832	28365618	21,38
rs916977	28513364	147,746
rs1667394	28530182	16,818

* Vzdálenost od předešlého markeru dle GRCh37.p5

Z porovnání 231 párů SNP odhalil test LD 27 případů statisticky významné vazebné nerovnováhy, z toho 15 bylo mezi markery SNP lokalizovanými na různých chromozomech nebo chromozomálních oblastech. Ovšem 12 případů vazebné nerovnováhy, které jsou uvedeny v tabulce 4.18, bylo mezi markery ležícími na 15. chromozómu v oblasti OCA2-HERC2. Tyto výsledky mohou být vysvětleny malou vzdáleností mezi těmito markery (viz Tabulka 4.19).

4.4.2 Asociační test

Výsledky asociačního testu byly rozdílné pro souhrnná data a pro bělošská data. U 3 markerů SNP byla zaznamenána statisticky významná asociace s barvou očí jak u souhrnných, tak i u bělošských vzorků: rs12913832, rs916977 a rs1667394 (viz Ta-

bulka 4.20 a Tabulka 4.21). Markery rs16891982 v genu *SLC45A2*, rs1426654 v genu *SLC24A5*, rs4778241 v genu *OCA2* a rs885479 v genu *MC1R* vykazovaly statisticky významnou asociaci s barvou očí jen u souhrnných vzorků. Hladina významnosti byla stanovena na 0,0023 po Bonferroniho korekci.

Tabulka 4.20: Výsledky asociační studie určování barvy očí provedené s použitím bělošských vzorků

Marker	MA	MAF(S)	MAF(T)	χ^2	p-hodnota	OR	D95	H95
rs16891982	C	0,07447	0,08491	0,07374	0,786	0,8672	0,3098	2,427
rs12203592	T	0,07447	0,08491	0,07374	0,786	0,8672	0,3098	2,427
rs35264875	T	0,1702	0,1887	0,1151	0,7344	0,8821	0,4271	1,822
rs1393350	A	0,2021	0,2075	0,008978	0,9245	0,9673	0,486	1,925
rs12821256	C	0,07447	0,1415	2,287	0,1304	0,4881	0,1899	1,255
rs2031526	A	0,234	0,1321	3,509	0,06102	2,008	0,9603	4,199
rs12896399	T	0,4043	0,3396	0,8928	0,3447	1,319	0,742	2,346
rs1800407	A	0,04255	0,1132	3,379	0,06603	0,3481	0,1083	1,119
rs7495174	G	0,117	0,2075	2,963	0,08517	0,506	0,2309	1,109
rs12913832	A	0,1489	0,5566	35,72	2,27E-09	0,1394	0,07027	0,2766
rs916977	A	0,06383	0,3302	21,69	3,21E-06	0,1383	0,05508	0,3473
rs1426654	G	0,01064	0,009434	0,007299	0,9319	1,129	0,06964	18,31
rs4778138	G	0,1702	0,2264	0,9835	0,3213	0,7009	0,3465	1,418
rs1545397	T	0,07447	0,06604	0,05439	0,8156	1,138	0,3839	3,373
rs4778241	A	0,1596	0,3302	7,735	0,005418	0,3852	0,1943	0,7637
rs1667394	G	0,07447	0,3396	20,75	5,22E-06	0,1564	0,06563	0,3729
rs258322	T	0,08511	0,08491	2,58E-05	0,9959	1,003	0,3704	2,713
rs1805007	T	0,04255	0,04717	0,02471	0,8751	0,8978	0,2339	3,446
rs885479	A	0,05319	0,06604	0,1458	0,7026	0,7945	0,2435	2,593
rs4911414	T	0,3404	0,3302	0,02344	0,8783	1,047	0,5815	1,885
rs1015362	A	0,2979	0,3019	0,003823	0,9507	0,9811	0,5351	1,799

statisticky významný výsledek ($p < 0,0023$), MA - minoritní alela, MAF(S) - frekvence minoritní alely u světlookých osob, MAF(T) - frekvence minoritní alely u tmavookých osob, OR - odds ratio, D95 - dolní mez 95% intervalu spolehlivosti, H95 - horní mez 95% intervalu spolehlivosti

Asociační test určování biogeografického původu odhalil statisticky významné

Tabulka 4.21: Výsledky asociační studie určování barvy očí provedené s použitím souhrnných vzorků

Marker	MA	MAF(S)	MAF(T)	χ^2	p-hodnota	OR	D95	H95
rs16891982	C	0,07447	0,2976	17,63	2,68E-05	0,1899	0,08214	0,439
rs12203592	T	0,07447	0,06548	0,07617	0,7826	1,148	0,4296	3,07
rs35264875	T	0,1702	0,1548	0,1069	0,7437	1,12	0,5668	2,214
rs1393350	A	0,2021	0,1488	1,226	0,2681	1,449	0,7499	2,8
rs12821256	C	0,07447	0,1071	0,7455	0,3879	0,6705	0,2693	1,669
rs2031526	A	0,234	0,2917	1,013	0,3142	0,7421	0,4147	1,328
rs12896399	T	0,4043	0,3095	2,399	0,1214	1,514	0,8946	2,561
rs1800407	A	0,04255	0,09524	2,373	0,1235	0,4222	0,1369	1,302
rs7495174	G	0,117	0,1845	2,04	0,1532	0,5857	0,2795	1,227
rs12913832	A	0,1489	0,6071	51,29	7,96E-13	0,1132	0,05931	0,2162
rs916977	A	0,06383	0,3869	31,85	1,67E-08	0,108	0,04466	0,2614
rs1426654	G	0,01064	0,2143	20,61	5,63E-06	0,03943	0,005311	0,2927
rs4778138	G	0,1702	0,2976	5,192	0,02269	0,4841	0,2575	0,9103
rs1545397	T	0,07447	0,2143	8,589	0,003382	0,295	0,1256	0,6928
rs4778241	A	0,1596	0,4167	18,18	2,01E-05	0,2658	0,1414	0,4999
rs1667394	G	0,07447	0,3988	31,29	2,23E-08	0,1213	0,05291	0,2781
rs258322	T	0,08511	0,2143	7,198	0,007298	0,3411	0,1513	0,7689
rs1805007	T	0,04255	0,02976	0,2973	0,5856	1,449	0,3795	5,532
rs885479	A	0,05319	0,2143	11,85	0,0005766	0,206	0,07783	0,5452
rs4911414	T	0,3404	0,2798	1,053	0,3048	1,329	0,7716	2,288
rs1015362	A	0,2979	0,2679	0,2702	0,6032	1,16	0,6633	2,027

statisticky významný výsledek ($p < 0,0023$), MA - minoritní alela, MAF(S) - frekvence minoritní alely u světlookých osob, MAF(T) - frekvence minoritní alely u tmavookých osob, OR - odds ratio, D95 - dolní mez 95% intervalu spolehlivosti, H95 - horní mez 95% intervalu spolehlivosti

odchylky ($\alpha = 0,0023$ po Bonferroniho korekci) u 10 markerů SNP (viz Tabulka 4.22). Část těchto markerů signifikantně asociuje s barvou očí (rs16891982, rs12913832, rs916977, rs1426654, rs4778241, rs1667394, rs885479). Ostatní markery (rs2031526, rs4778138, rs1545397, rs258322) neasociovaly statisticky signifikantně s barvou očí.

Tabulka 4.22: Výsledky asociační studie určování biogeografického původu provedené s použitím souhrnných vzorků

Marker	MA	MAF(B)	MAF(A)	χ^2	p-hodnota	OR	D95	H95
rs16891982	C	0,08	0,6613	93,95	3,24E-22	0,04454	0,02139	0,09272
rs12203592	T	0,08	0,03226	1,686	0,1941	2,609	0,5829	11,67
rs35264875	T	0,18	0,09677	2,435	0,1186	2,049	0,8198	5,12
rs1393350	A	0,205	0,04839	8,307	0,003948	5,071	1,513	17
rs12821256	C	0,11	0,04839	2,082	0,1491	2,431	0,7022	8,414
rs2031526	A	0,18	0,5645	35,42	2,66E-09	0,1693	0,09126	0,3142
rs12896399	T	0,37	0,2581	2,63	0,1049	1,688	0,8928	3,193
rs1800407	A	0,08	0,06452	0,1609	0,6883	1,261	0,4054	3,922
rs7495174	G	0,165	0,1452	0,1384	0,7099	1,164	0,5233	2,588
rs12913832	A	0,365	0,6935	20,71	5,35E-06	0,254	0,1377	0,4683
rs916977	A	0,205	0,4839	18,63	1,59E-05	0,2751	0,1502	0,5037
rs1426654	G	0,01	0,5645	120	6,34E-28	0,007792	0,001773	0,03425
rs4778138	G	0,2	0,4194	12,08	0,0005085	0,3462	0,1877	0,6383
rs1545397	T	0,07	0,4677	54,58	1,49E-13	0,08565	0,04097	0,1791
rs4778241	A	0,25	0,5645	21,36	3,81E-06	0,2571	0,1418	0,4664
rs1667394	G	0,215	0,5	18,97	1,33E-05	0,2739	0,1501	0,4997
rs258322	T	0,085	0,4355	41,61	1,12E-10	0,1204	0,05942	0,2441
rs1805007	T	0,045	0	2,889	0,08917	-	-	-
rs885479	A	0,06	0,4677	59,61	1,16E-14	0,07263	0,03371	0,1565
rs4911414	T	0,335	0,1935	4,496	0,03397	2,099	1,047	4,206
rs1015362	A	0,3	0,2097	1,921	0,1657	1,615	0,8166	3,196

statisticky významný výsledek ($p < 0,0023$), MA - minoritní alela, MAF(B) - frekvence minoritní alely u bělochů, MAF(A) - frekvence minoritní alely u asiátů, OR - odds ratio, D95 - dolní mez 95% intervalu spolehlivosti, H95 - horní mez 95% intervalu spolehlivosti

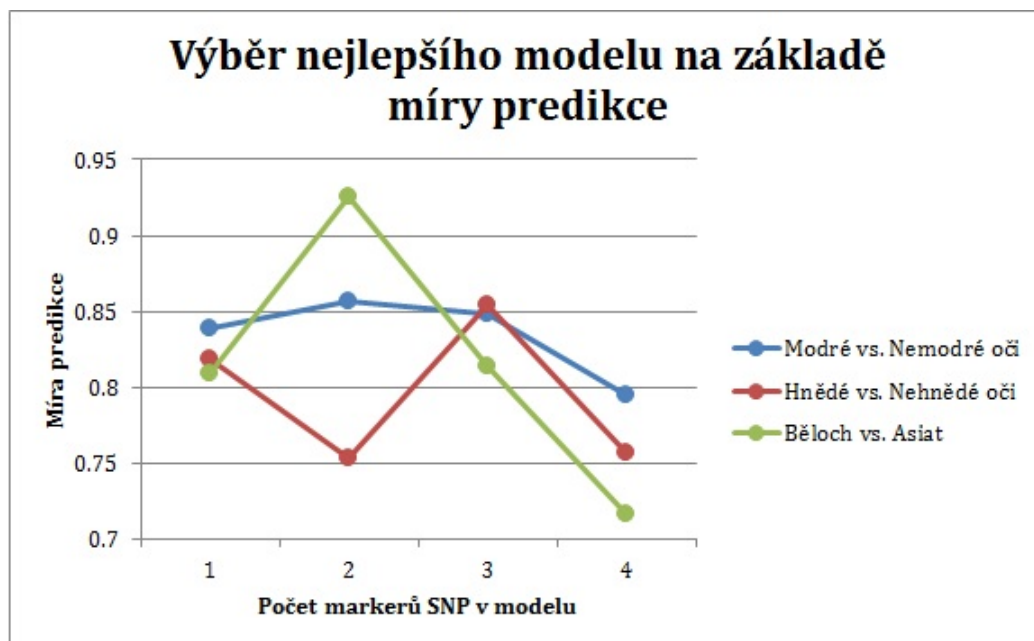
4.4.3 Analýza MDR

Před vlastní analýzou MDR bylo provedeno filtrování markerů SNP pomocí algoritmu ReliefF. Soubor následně vybraných markerů se téměř shodoval pro souhrnné a bělošské vzorky (viz Tabulka 4.23). Společně vybrané markery byly rs12913832, rs916977, rs1667394 a rs4778241. Rozdíl byl pouze v jednom markeru: u bělošských vzorků byl vybrán marker rs7495174, u souhrnných vzorků - rs16891982. Marker, který byl vybrán u souhrnných vzorků, je asociován s biogeografickým původem. Zajímavé je, že u vzorků rozdělených dle biogeografického původu byl vybrán polymorfismus rs12913832, který je spojen s barvou očí.

Tabulka 4.23: Výsledky filtrování ReliefF

Světlé vs. tmavé oči				Běloch vs. asiat	
Souhrnné vzorky		Bělošské vzorky			
SNP	ReliefF	SNP	ReliefF	SNP	ReliefF
rs12913832	0,386	rs12913832	0,298	rs16891982	0,787
rs916977	0,269	rs916977	0,222	rs2031526	0,568
rs1667394	0,219	rs1667394	0,184	rs258322	0,518
rs16891982	0,146	rs4778241	0,104	rs1426654	0,481
rs4778241	0,118	rs7495174	0,068	rs12913832	0,406

Nejlepší modely pro určování barvy očí a biogeografického původu byly vybrány na základě nejvyšší míry predikce a největší hodnoty CV (viz Obrázek 4.2). Nejlepší modely v této studii měly hodnoty CV 10/10. Výsledky analýzy MDR, stejně jako asociačního testu, se lišily v závislosti na použitém souboru vzorku (viz Tabulka 4.24). Nehledě na tuto skutečnost zahrnovaly nejlepší modely určování barvy očí pro oba soubory vzorků marker rs12913832. Míra predikce byla 0,8566 pro souhrnné vzorky a 0,8549 pro bělošské vzorky. Na obrázku 4.3 je zobrazen výstup z programu MDR pro nejlepší model pro souhrnné vzorky obsahoval 2 SNP: rs12913832 a rs916977. Nejlepší model pro bělošské vzorky obsahoval navíc jeden marker: rs7495174. Největší míra predikce byla u modelu určování biogeografického původu, který obsahoval markery rs1426654 a rs16891982. Tato míra predikce byla 0,9255.



Obrázek 4.2: Výběr nejlepšího modelu na základě míry predikce

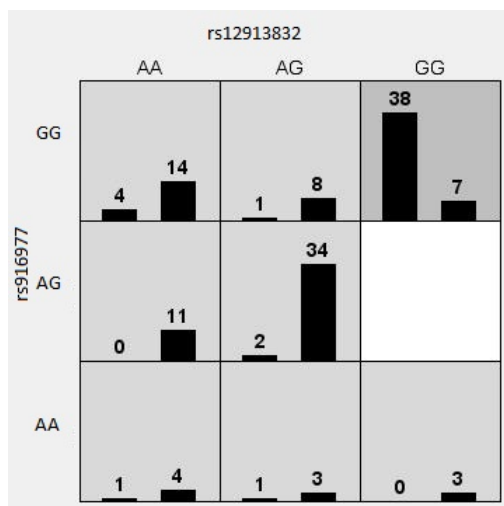
Tabulka 4.24: Výsledky analýzy MDR

Světlé vs. tmavé oči						Běloch vs. asiat		
Souhrnné vzorky			Bělošské vzorky					
SNP	Míra predikce	CV	SNP	Míra predikce	CV	SNP	Míra predikce	CV
rs12913832	0,8388	10/10	rs12913832	0,8193	10/10	rs1426654	0,8098	6/10
rs12913832 rs916977	0,8566	10/10	rs12913832 rs916977	0,7533	7/10	rs16891982 rs1426654	0,9255	10/10
rs16891982 rs1291383 rs1667394	0,8490	10/10	rs7495174 rs12913832 rs916977	0,8549	10/10	rs16891982 rs2031526 rs1426654	0,8137	6/10
rs16891982 rs1291383 rs916977 rs4778241	0,7955	8/10	rs7495174 rs12913832 rs916977 rs4778241	0,7569	10/10	rs16891982 rs2031526 rs1291383 rs1426654	0,7169	10/10

statisticky významný výsledek ($p < 0,0023$), CV - cross validation

4.4.4 Rekonstrukce haplotypů

Pomocí analýzy MDR byly vybrány 3 markery pro určování barvy očí, které se nacházejí na 15. chromozómu v oblasti genů HERC2-OCA2. Proto byla pro tyto tři markery provedena rekonstrukce haplotypů. V tabulce 4.25 jsou uvedeny výsledky



Obrázek 4.3: Výstup analýzy MDR pro určení barvy očí u souhrnných vzorku

V každé buňce představuje první sloupec počet světlookých osob a druhý sloupec počet tmavookých osob

rekonstrukce haplotypů rozdělené dle barvy očí osob. Nejčastější haplotyp u obou skupin byl AGG, ovšem u světlookých lidí byla četnost tohoto haplotypu 84,04%, zatímco u tmavookých lidí 33,93%.

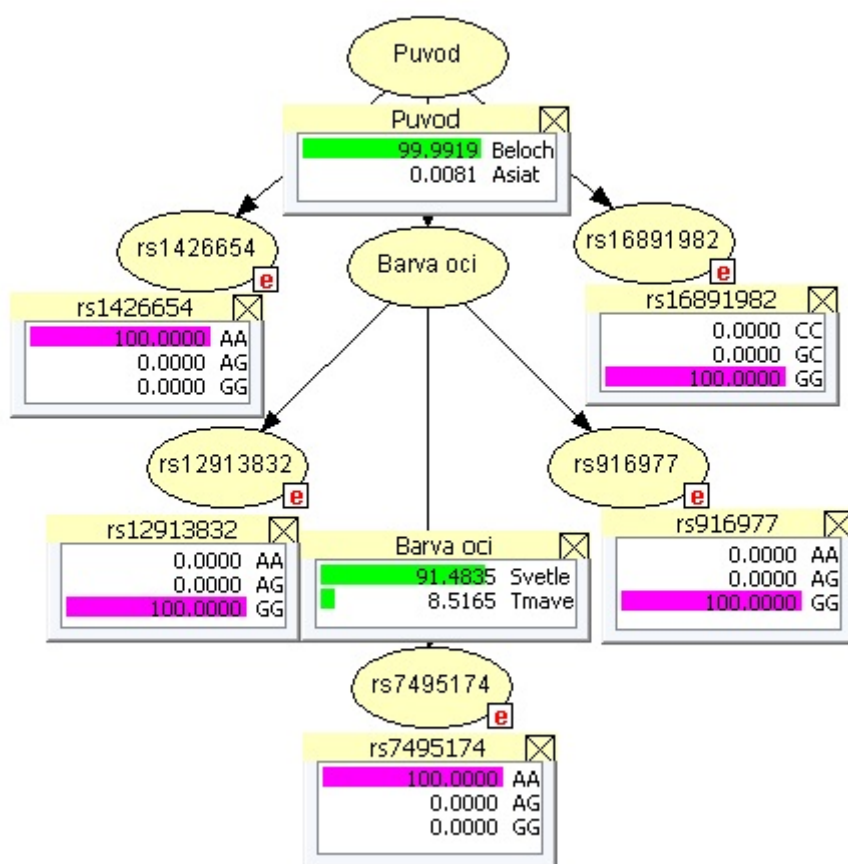
Tabulka 4.25: Výsledky analýzy MDR

Haplotyp*	Frek. (S)	Frek.(T)
AAA	0,0213	0,2679
AAG	0,0106	0,1548
AGA	0,0106	0,0536
AGG	0,8404	0,3393
GAA	0,0319	0,0655
GAG	0,0851	0,119

**rs7495174, rs12913832 a rs916977, Frek.(S) - frekvence u světlookých osob, Frek.(T) - frekvence u tmavookých osob*

4.4.5 Vytvoření Bayesovy sítě

Byly vytvořeny 2 modely Bayesových sítí (BS) pro zjištění, jak implementace biogeografického původu ovlivní parametry BS. První BS (BS1) obsahovala markery, které byly vybrány pomocí MDR pro určování barvy očí (rs7495174, rs12913832 a rs916977). Rodičovský uzel „Barva očí“ měl 3 dětské uzly: „rs7495174“, „rs12913832“ a „rs916977“. Druhá BS (BS2), která je znázorněna na obrázku 4.4, byla vytvořena kombinací nejlepších modelů pro určování barvy očí a biogeografického původu vybraných pomocí analýzy MDR a obsahovala následující markery: rs7495174, rs12913832, rs916977, rs1426654 a rs16891982. Rodičovský uzel „Původ“ měl 3 dětské uzly: „rs1426654“, „rs16891982“ a „Barva očí“. Poslední uzel měl, stejně jako v případě první BS, 3 dětské uzly: „rs7495174“, „rs12913832“ a „rs916977“.



Obrázek 4.4: Vytvořena Bayesova síť pro určování barvy očí a biogeografického původu

4.4.6 Validace Bayesovy sítě

Validace vytvořené BS byla provedena s použitím jak bělošských, tak i asijských vzorků. Pomocí těchto vzorků byly stanoveny důležité parametry úspěšnosti pro obě BS (viz tabulka 4.26). U obou BS byla přesnost určování barvy očí více než 98%. Senzitivita a specifita pro určování barvy očí dosahovaly u obou BS 96% a 98%. Během validace byl biogeografický původ určen správně u všech testovaných osob (N=128). Proto všechny parametry této BS pro určování biogeografického původu dosáhly nejvyšší možné hodnoty - 1,0. Míra neprůkazných výsledků byla menší pro BS2 (0,016), než pro BS1 (0,133). Během určování biogeografického původu nebyl zaznamenán žádný neprůkazný výsledek.

Tabulka 4.26: Charakteristiky vytvořené Bayesovy sítě pro určování barvy očí a biogeografického původu

	Světlé vs. tmavé oči		Běloch vs. asiat
	BS 1	BS 2	BS 2
Senzitivita	0,9600	0,9600	1,0000
Specifita	0,9881	0,9899	1,0000
PPH	0,9600	0,9600	1,0000
NPH	0,9881	0,9899	1,0000
Přesnost	0,9817	0,9839	1,0000

BS - Bayesova síť, BS 1 - rs7495174, rs12913832, rs916977, BS 2 - rs16891982, rs1426654, rs7495174, rs12913832, rs916977, PPH - pozitivní prediktivní hodnota, NPH - negativní prediktivní hodnota

5 DISKUZE

5.1 Investigator DIPplex

V této části studie byla testována sada markerů Investigator DIPplex s cílem zjistit, zda je tato sada vhodná pro použití v české populaci pro forenzní a paternitní účely. Bylo zjištěno, že sada DIPplex je velmi citlivá a vyžaduje malé vstupní množství DNA (500 pg). Navíc vyšší vstupní množství DNA způsobuje vytažené a abnormálně tvarované píky, což bylo doloženo také studiemi dalších vědeckých skupin [116]. Vytažení píků je dáno designem sady Investigator DIPplex, kde dochází k velikostnímu překryvu alel různých polymorfismů značených různými fluorescenčními barvami. Tato skutečnost pak způsobuje, že pokud je osoba homozygot v některé z těchto alel, výsledkem analýzy jsou píky vytažené ve všech barevných kanálech.

Ani jeden ze 30 markerů INDEL obsažených v sadě DIPplex nevykazoval odchylku od Hardy-Weinbergovy rovnováhy, což dovoluje použít zákony Hardyho-Weinberga pro odhad frekvencí genotypů v české populaci. Vazebná rovnováha zkoumaných markerů svědčí o jejich nezávislosti. Toto zjištění umožňuje použít takzvané „produktové pravidlo“, dle kterého se průnik pravděpodobností nezávislých jevů (v tomto případě to jsou genotypy v jednotlivých lokusech) rovná součinu pravděpodobností těchto jevů. Průměrná heterozygotita zkoumaných markerů v české populaci byla 0,46, tedy relativně vysoká, což je velmi žádoucí pro identifikační markery. Důvodem je to, že větší heterozygotita zvyšuje pravděpodobnost toho, že se genotypy osob ve zkoumaných markerech budou lišit. Téměř 50% průměrná heterozygotita svědčí o tom, že jsou zkoumané lokusy v české populaci vysoce polymorfní. Hodnota heterozygotity byla srovnatelná s hodnotami heterozygotity u bělošských populací: dánská populace - 0,49 [116], středošpanělská - 0,49 [117], finská - 0,47 [118], korejská - 0,40 [119]. Největší rozdíl v průměrné heterozygotitě byl u korejské populace, která zároveň vykazuje její nejnižší hodnotu, což může být způsobeno tím, že sada DIPplex byla vyvinuta na populaci německé.

Forenzní parametry, které byly zjištěny v této studii, dovolují hodnotit užitečnost zkoumaných markerů jak jednotlivě, tak i v sadě. Jeden z nejdůležitějších parametrů, diskriminační síla, nabývala hodnot pro českou populaci v rozsahu od 0,554 do 0,656. Rozsah PD byl téměř stejný jako ve finské populaci (0,548 - 0,656) [118]. Pro španělskou populaci byla nejvyšší hodnota PD rovná 0,645 [117]. Hodnoty

PD v korejské populaci vykazovaly větší rozsah, od 0,211 do 0,647 [119], což je nejpravděpodobněji způsobeno výše zmíněnou nižší heterozygotitou v některých markerech. CPD pro celou sadu markerů byla u české populace 99,999999999 %, což je obdobné jako u ostatních studovaných populací. Pravděpodobnost náhodné shody, která souvisí s diskriminační silou, byla u české populace 1 z $6,8 \times 10^{12}$, což je nižší hodnota než u rutinně používaných sad STR markerů, ovšem tato hodnota je srovnatelná s ostatními studiemi sad markerů INDEL. Průměrný paternitní index v každém ze zkoumaných markerů přesahoval 1, což naznačuje relativní užitečnost pro testování paternity. API markerů INDEL byl zhruba o třetinu menší než API markerů STR [118]. Ovšem kombinovaný paternitní index pro všech 30 markerů byl kolem 27 000, takže výsledná průměrná pravděpodobnost otcovství byla 99,99% za předpokladu 50% výchozí pravděpodobnosti. I když PI pro sadu DIPplex nedosahuje hodnot PI pro STR lokusy, je pořád dostačující pro prakticky prokázané otcovství. Sada DIPplex může také sloužit jako doplňující sada markerů. Například po doplnění analýzy pomocí Powerplex 16 HS sadou DIPplex vzroste výsledná pravděpodobnost otcovství z 99,99999 na 99,999999999. Sada markerů INDEL může být obzvlášť užitečná v komplikovaných případech určování otcovství, z důvodu nízké mutační rychlosti těchto polymorfizmů. S problémem určování otcovství je také spojena determinace maximálně informativních markerů. V této studii byl počet účinných markerů stanoven na 29,76, což je mírně vyšší hodnota, než ve finské populaci - 28,32 [118].

Test porovnání alelických frekvencí odhalil nejvíce statisticky významných rozdílů mezi českou a indiánskou populací. Žádné statisticky významné rozdíly nebyly nalezeny mezi českou a německou populací. Tyto výsledky jsou podporovány skutečností, že německá populace byla nejbližší české populaci ze všech testovaných populací. Relativně malý počet markerů, u nichž byly nalezeny statisticky významné odchylky, svědčí o robustnosti zkoumané sady markerů. Toto je velmi důležité ve forenzní genetice, například pro identifikační analýzu vzorku z místa činu, který zanechal pachatel neznámé národnosti. Pravděpodobnost shody u robustní sady markerů by neměla významně kolísat v závislosti na populaci, čímž se zvýší hodnota výsledků analýzy pomocí této sady markerů.

Výsledky populační studie Investigator DIPplex v České republice byly publikovány v časopise *International Journal of Legal Medicine* [120].

5.2 Mentype[®] Argus X-8

Tato studie se zabývala populačním výzkumem sady markerů Mentype[®] Argus X-8, která se skládá ze 4 vazebných skupin obsahující dvojice markerů. Markery ve dvojici jsou lokalizovány na chromozómu X v těsné blízkosti.

Všechny zkoumané markery ze sady Argus X-8 byly v HWE, což je jedním z kritérií pro použití mikrosatelitních markerů ve forenzně-genetické praxi [121]. Dalším kritériem je informace o vazebné nerovnováze. I přesto, že markery jsou ve vazebných skupinách blízko u sebe, nebyl nalezen žádný pár markerů, jejichž výskyt by ve studovaných vzorcích vykazoval vazebnou nerovnováhu. Vysvětlení těchto výsledků může být několik: malý počet zkoumaných vzorků, anebo vysoká mutační rychlost zkoumaných markerů STR [26]. Stejně výsledky vazebné nerovnováhy byly pozorovány také ve studiích na jiných populacích [122], [123].

Zkoumané markery se ukázaly jako vysoce polymorfní ve středochorvatské populaci. Největší počet alel měl marker DXS10135 - 25, nejmenší počet měly DXS8378 a DXS7423 - 5.

Nejužitečnější marker z celé sady Argus X-8 byl DXS10135, který měl nejvyšší hodnoty informačního obsahu ($PIC=0,9306$) a diskriminační síly u žen ($PD_F=0,9918$) i u mužů ($PD_M=0,9345$). Byl to také marker s nejvyšší pozorovanou heterozygotitou ($0,9596$) a s největším počtem alel. Výsledky této studie se shodovaly s výsledky jiných populačních studií, v nichž byl marker DXS10135 nejlepším ze všech zkoumaných markerů ze sady Argus X-8 [122], [109]. Nejméně informativní marker byl ve středo-chorvatské populaci DXS8378, jehož hodnota informačního obsahu byla $0,6316$ a diskriminační síla byla pro ženy $0,8447$ a pro muže $0,6922$. DXS8378 vykazoval nejmenší pozorovanou heterozygotitu - $0,6768$. Ve studiích sady markerů Argus X-8 ostatních vědeckých skupin byl nejméně informativní buď stejný marker jako v této studii - DXS8378 [110], [124] nebo DXS7423 [122], [109]. Marker DXS7423 byl v této studii druhým nejméně informativním. Získané výsledky naznačují, že vytvořená sada markerů X-STR je vhodná pro univerzální použití v nezávislosti na původu osoby.

Velmi důležitým parametrem byla pro markery lokalizované na chromozómu X průměrná síla vyloučení v případech určování otcovství. Důvodem je velká informativita těchto markerů nejen pro určování otcovství dítěte ženského pohlaví, ale i v případech určování sourozenectví a neúplného sourozenectví, kde nejsou au-

tozomální markery až tak užitečné. Tuto skutečnost podporuje zjištění, že pokud mají autozomální markery a markery na chromozómu X stejnou heterozygotitu a hodnotu informačního obsahu, potom markery umístěné na chromozómu X mají větší průměrnou sílu vyloučení otcovství [25]. Hodnoty průměrné síly vyloučení pro různé případy dovolují znázornit, jak je zkoumaný marker užitečný pro určování příbuznosti. Velmi zajímavým parametrem je průměrná síla vyloučení například v případě paternitního tria, kde je místo nařčeného otce vyšetřována jeho matka, tedy údajná babička dítěte (MEC pro deficientní případy). Nejvyšší hodnota tohoto parametru byla u již zmíněného markeru DXS10135 - 86,7%, nejnižší u DXS8378 - 42%. Jinými slovy, pokud nařčený otec není otcem, je při použité markeru DXS10135 více než 85% pravděpodobnost, že bude otcovství v tomto markeru vyloučeno. V případě určování otcovství u dua nebo tria je k dipozici více informací, a proto jsou tyto hodnoty vyšší, než výše zmíněný parametr, což je podloženo výsledky této studie. Například pro marker DXS10135 je hodnota síly vyloučení v deficientním případě (babička místo nařčeného otce) 86,7%, v případě dua 87,4% a v případě tria 93%. Pokud se jedná o případ s počtem meióz větší než jedna, jako například při určování úplného a neúplného sourozenectví, je velmi vhodné použít pro výpočet příbuznosti místo jednotlivých lokusů haplotypy.

Protože jsou zkoumané markery umístěné ve vazebných skupinách a jsou vhodné pro určování příbuznosti, byly vypočítány frekvence haplotypů pro každou vazebnou skupinu. Pozorovaný počet haplotypů u LG1 (DXS10135 - DXS8378), LG2 (DXS7132 - DXS10074), LG3 (HPRTB - DXS10101) a LG4 (DXS10134 - DXS7423) byl 37, 30, 35, 30. Počet možných haplotypů pro tyto vazebné skupiny byl 3-4 krát vyšší než pozorovaný počet: 125, 91, 162 a 115. Nejčastější haplotypy byly v LG1 25-12(8%) a 24-12 (8%), v LG2 13-16 (11,5%), v LG3 13-31.2 (10%) a v LG4 35-13 (10%). Relativní homogenita haplotypů u všech vazebných skupin svědčí o vazbě. Používání haplotypové analýzy na X-chromozómu by mohlo pomoci nejen v určování příbuznosti, ale také pro určení biogeografického původu, pokud bude shromážděn dostatek dat (haplotypových frekvencí) u ostatních světových populací. Proto bylo provedeno srovnání haplotypových frekvencí, které odhalilo statisticky významné rozdíly nejen mezi středo-chorvatskou, japonskou populací a populací Ghany, ale také mezi středo-chorvatskou a německou populací, které jsou si mnohem blíží.

Protože ne všechny studie určily haplotypové frekvence, bylo také provedeno

porovnání genetické vzdáleností populací pomocí alelických frekvencí. Byly vybrány dvě blízké populace - polská, u které byly známy alelické frekvence pro všech 8 markerů, a bosenská, která je velmi blízká chorvatské populaci a u které byly známy frekvence 4 markerů. Porovnání genetických vzdáleností neodhalilo žádnou statisticky významnou odchylku, což lze vysvětlit výše zmíněnou příbuzností porovnávaných populací.

5.3 Decaplex X-STR

Sada markerů Decaplex na rozdíl od Mentype[®] Argus X-8 obsahovala markery na chromozómu X, které nejsou ve vazebných skupinách (kromě jedné dvojice markerů - viz dále). Všechny zkoumané markery byly v HWE, což je, jak již bylo zmíněno, nezbytným předpokladem pro forenzní markery [121]. Testování vazebné nerovnováhy mělo u této sady markerů v české populaci rozdílné výsledky u mužských a ženských vzorků. Zatímco u mužských vzorků po testování všech párů lokusů nebyla zjištěna vazebná nerovnováha ani u jednoho páru, u ženských vzorků byla zjištěna vazebná nerovnováha mezi markery DXS6809 a DXS6789. Všechny markery ze sady Decaplex mají vzdálenost větší než 5 Mb [30] kromě již zmíněných DXS6809 a DXS6789, u kterých byla pozorována vazebná nerovnováha. Vzdálenost mezi touto dvojicí markerů je kolem 500 Kb, takže zjištěná vazebná nerovnováha je nejen statisticky významná, ale je podložena výsledky lokalizace markerů STR na chromozómu X. Na druhou stranu jsou výsledky vazebné nerovnováhy markerů na chromozómu X u mužských vzorků více spolehlivé než výsledky vazebné nerovnováhy u ženských vzorků. Důvodem je proces testování vazebné nerovnováhy. Například dva markery jsou ve vazebné rovnováze, pokud jsou frekvence výskytu alely A jednoho markeru a alely B druhého markeru na sobě nezávislé. Z toho následně vyplývá, že se frekvence haplotypu AB rovná součinu frekvencí alel A a B. Proto je nezbytné znát haplotypy testovaných vzorků, aby bylo možné provést testování vazebné nerovnováhy. U mužských vzorků jsou haplotypy markerů na chromozómu X známy, protože muži mají jenom jeden chromozóm X, u ženských vzorků je potřeba nejprve provést rekonstrukci haplotypů, která má určitou míru přesnosti, což se především týká rekonstrukce vzácných haplotypů. Pokud se jedná o příbuzenskou analýzu, je dle doporučení Mezinárodní společnosti pro forenzní genetiku (International Society for Forensic Genetics - ISFG) potřeba počítat s frekvencí haplotypů pro markery

DXS6809 a DXS6789 [30]. Z teoreticky možných 144 haplotypů bylo pozorováno 52 u 234 mužů, přičemž nejčastější haplotypy byly 31-22 a 34-22. Pokud se jedná o identifikaci, mohou být frekvence haplotypů odhadnuty pomocí frekvencí jednotlivých alel [30].

Nejvíce alel bylo nalezeno u polymorfních markerů DXS6809 a DXS6789 (12 alel), nejméně u DXS7133, DXS7423 a DXS8378 (5 alel). Dva poslední markery byly nejméně informativní také u středochovatské populace. Nejinformativnější byl marker DXS6809, jehož diskriminační síla byla pro muže 82% a pro ženy 95%. Zajímavé je, že DXS6789 nebyl tak informativní, pravděpodobně proto, že obsahoval alelu, která měla frekvenci 41%, a další dvě alely s frekvencí 24% a 16%. Nejméně informativní byl marker DXS7133: PD_F zde dosahovala hodnoty 81 %, PD_M 64%.

Porovnání genetické vzdálenosti pomocí alelických frekvencí mezi českou, ugandskou, severovýchodní španělskou, německou a polskou populací odpovídalo geografickému rozložení populací. Česká populace se nejvíce lišila od populace Ugandy, severovýchodního Španělska a Německa. Porovnání s nejbližší, tedy polskou populací nezjistilo žádné statisticky významné odchylky.

Dvě výše popsané studie markerů lokalizovaných na chromozómu X Mentype® Argus X-8 a Decaplex X-STR představují dvě možné varianty navržení sady X-chromozomálních markerů. Obě možnosti mají svoje výhody a nevýhody. Použití vazebných skupin markerů je méně informativní v případě identifikace osob, ovšem je velmi vhodné pro určování příbuznosti, obzvlášť u komplexních nebo deficientních případů. Sada Decaplex X-STR obsahující nezávislé markery je vhodnější pro účely identifikace. Pokud se použije pro určování příbuznosti, nenastává problém nulových frekvencí, který se často objevuje u haplotypové analýzy. Nulové frekvence jsou velmi často reportovány u vzácnějších haplotypů, protože populační studie zkoumající haplotypy tvořené více markery potřebuje mnohem větší počet vzorků než populační studie zkoumající jednotlivé markery.

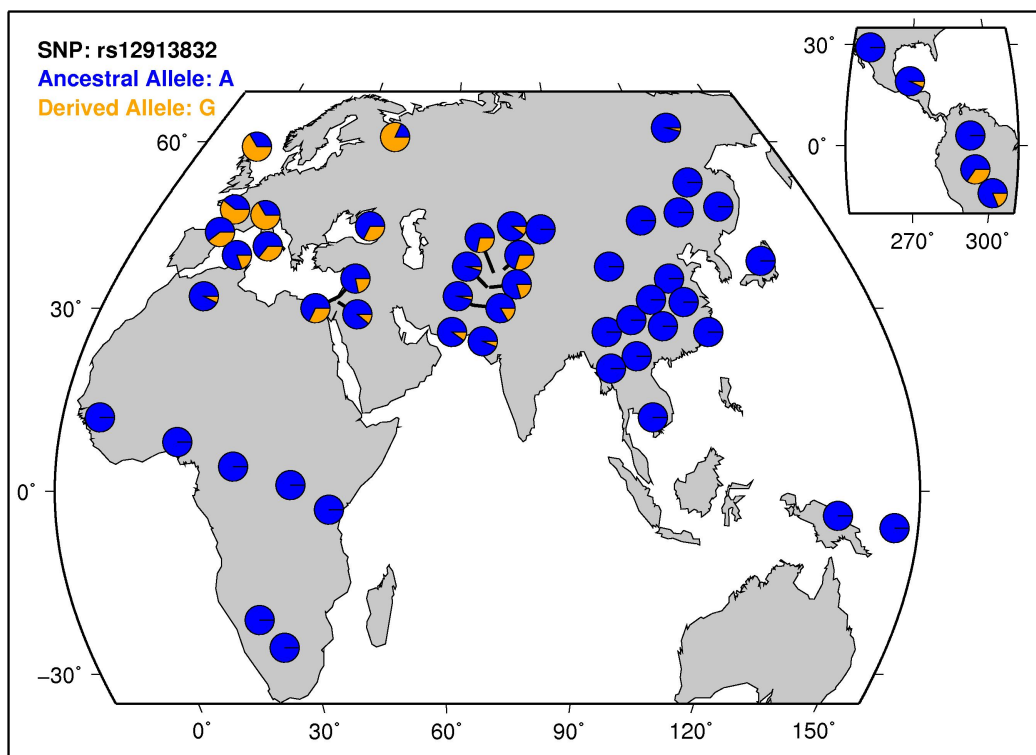
Výsledky populační studie Decaplex X-STR v České republice byly publikovány v časopise International Journal of Legal Medicine [125].

5.4 Fenotypové SNP

Tato studie zjistila, že marker rs6119471 není polymorfní ve zkoumaném vzorku bělošské a asijské populace. Toto zjištění se shoduje s výsledky ostatních studií, kde

byl tento polymorfismus zjištěn jenom u africké populace [70]. Provedená studie prokázala, že marker rs12913832 je velmi silně asociován s barvou očí, ale může být také použit jako marker biogeografického původu. Na obrázku 5.1 jsou znázorněny frekvence alel markeru rs12913832. Lze říci, že oblasti výskytu odvozené alely G se shodují s oblastmi výskytu modré barvy očí. V této studii mělo v bělošských vzorcích 20 osob v markeru rs12913832 genotyp AA, 33 genotyp AG a 47 genotyp GG, který byl majoritní. Asijské vzorky vykazovaly opačnou tendenci: 14 osob mělo genotyp AA, 15 osob genotyp AG a jenom 2 osoby genotyp GG. Výsledky asociační studie u souhrnných vzorků bělochů a asiátů se shodovaly ve třech markerech, ale měly navíc další 4 statisticky významné markery, které asociovaly s barvou očí: rs1426654, rs16891982, rs4778241 a rs885479. Tato studie také potvrdila zjištění předešlých studií [126], že SNP rs1426654 a rs16891982 jsou vhodné markery pro určování biogeografického původu (viz Obrázek 5.2 a Obrázek 5.3). Ani jedna osoba ze 100 zkoumaných bělochů neměla genotyp rs1426654 GG, zatímco z 31 zkoumaných asiátů byl tento genotyp zjištěn u 11 osob. Pro marker rs16891982 nebyl nalezen mezi bělošskými vzorky žádný homozygot CC, zatímco u asijských vzorků mělo genotyp CC 13 osob. Zajímavé je, že oba výše uvedené markery také asociovaly s barvou očí, konkrétně alela G u rs1426654 a alely C rs16891982 s tmavou barvou očí. To vede k domněnce, že se silná asociace s nebělošským biogeografickým původem promítá i do asociace s tmavou barvou očí. Předešlá studie vědecké skupiny Rana [127] zjistila, že polymorfismus rs885479 v genu MC1R téměř není přítomen u bělochů, ale je naopak častý u asiátů. V naší studii neměl žádný z bělochů v rs885479 genotyp AA, zatímco 7 osob asijského původu mělo genotyp AA. Obdobné frekvence genotypů byly pozorovány pro marker rs1545397, který je asijskou specifickou mutací v genu OCA2 [82]. Tento marker sice nevykazoval statisticky významnou asociaci, ale jeho p-hodnota byla relativně nízká ($p=0,0034$, $\alpha=0,0023$). Frekvence alel polymorfismu rs4778241 se lišila mezi bělošskými a asijskými osobami, což bylo pozorováno i v jiných studiích [128].

Použití analýzy MDR odhalilo polymorfizmy, které by mohly být využity k optimalizaci predikce barvy očí, ovšem neměly statisticky významnou asociaci s barvou očí během asociačního testu provedeného u jednotlivých markerů. Zajímavé je, že jako nejinformativnější marker pro určování barvy očí a biogeografického původu byl potvrzen marker rs12913832. Tento marker byl vybrán jako nejlepší model obsa-

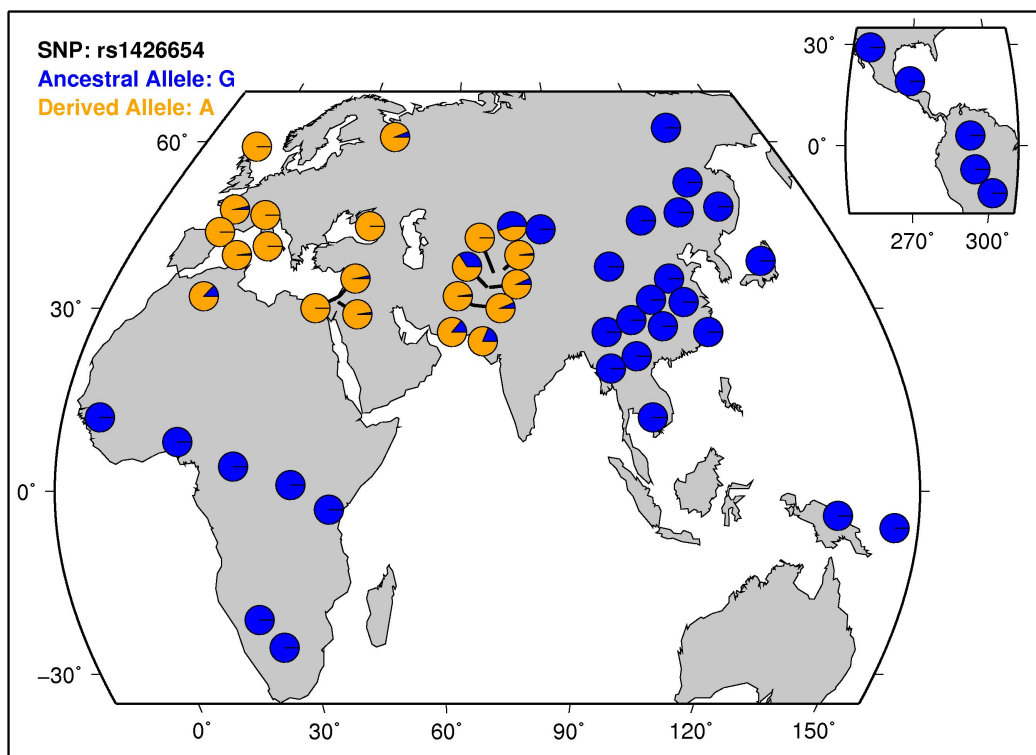


Obrázek 5.1: Frekvence alel SNP rs12913832 u světových populací

Zdroj: <http://hgdp.uchicago.edu/>

hující jeden SNP ze všech markerů po filtrování ReliefF. Stejný výsledek byl získán ve studii Pospiech et al. [71] pro určování modré vs. nemodré barvy očí a hnědé vs. nehnědé barvy očí, což potvrzuje predikční účinnost markeru rs12913832. Ve studii Ruiz et al. [69], byla po provedení analýzy MDR pro určování hnědé vs. nehnědé barvy očí vybrána jako nejlepší kombinace rs12913832 a rs4778138. V téže studii byla analýza MDR také použita pro určování zelenohnědých vs. nezelenohnědých očí, v níž byl vybrán jako nejlepší model kombinace SNP obsahující rs12913832 a rs1667394. Markery rs4778138 a rs1667394 byly také testovány v této studii, ovšem nebyla u nich pozorována asociace s barvou očí na zkoumaných vzorcích. Důvodem může být populační specificita těchto dvou markerů. Míra predikce byla v této studii pro určování světlé vs. tmavé barvy u bělošských vzorků a ve studii Pospiech et al. [71] pro určování modré vs. nemodré barvy podobná u obou studií (0,8369 a 0,8193). Obdobné výsledky hodnot míry predikce byly dosaženy i ve studii Ruiz et al. [69]: hnědé vs. nehnědé - 0,8587 a zelenohnědé vs. nezelenohnědé - 0,8322.

Velmi zajímavými se ukázaly být výsledky rekonstrukcí haplotypů pro markery rs7495174, rs12913832 a rs916977. Z 8 možných haplotypů bylo jak u tmavookých

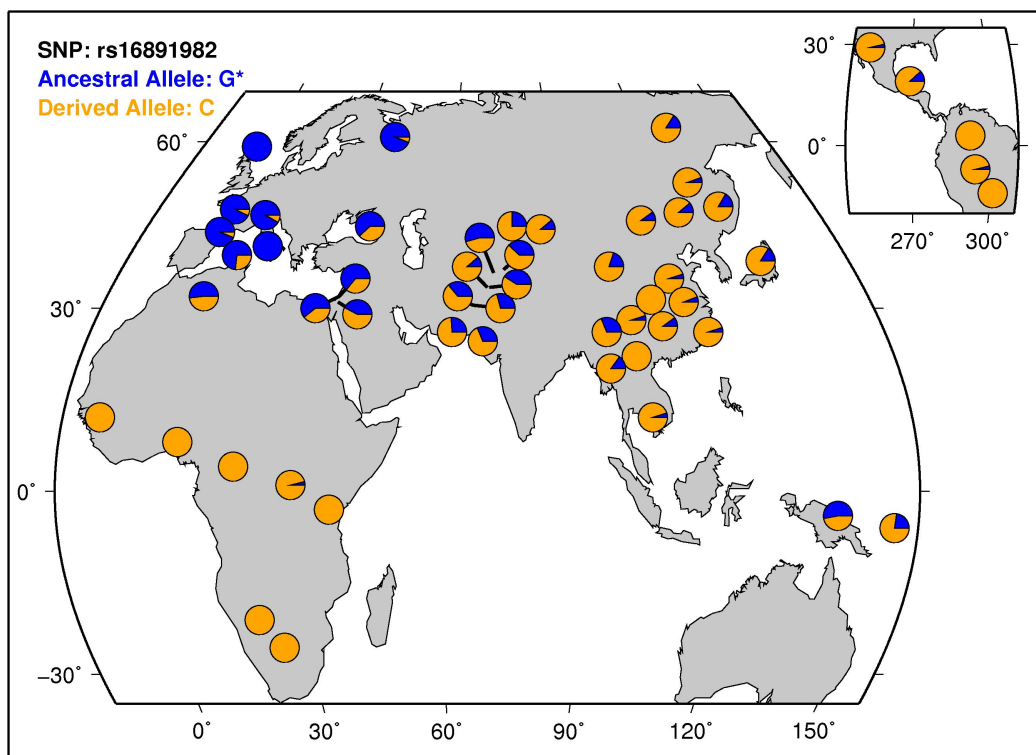


Obrázek 5.2: Frekvence alel SNP rs1426654 u světových populací

Zdroj: <http://hgdp.uchicago.edu/>

osob, tak u světlookých osob pozorováno 6 haplotypů. U žádné osoby nebyl nalezen haplotyp GGG nebo GGA. Kromě vzácnosti těchto haplotypů, může být důvodem jejich nenalezení v naší studii i skutečnost, že na základě dat získaných ze vzorků byly haplotypy rekonstruovány pomocí algoritmu, který má určitou přesnost. Světlooké osoby byly relativně homogenní co se týče haplotypů. Frekvence majoritního haplotypu (AGG) v této skupině dosahovala 84% a frekvence ostatních haplotypů se pohybovala pod 10%. Haplotypy tmavookých osob zdaleka nebyly tak homogenní. Byly zde nalezeny 4 majoritní haplotypy: AGG (34%), AAA(27%), AAG (16%) a GAG (12%). Jedním z vysvětlení mohou být různé subfenotypy (od zelenohnědé do černohnědé barvy očí), které jsou součástí skupiny tmavých očí. Světloocí jedinci měli v 81% diplotyp AGG/AGG, zatímco tmavoocí jedinci měli nejčastěji haplotyp AGG jenom na jednom chromozómu. Tyto výsledky podporují již dříve formulovanou hypotézu o tom, že modrá barva očí je nejčastěji děděna jako recesivní znak [129].

Výhodou analýzy MDR je tvorba třídícího algoritmu, který přisuzuje určitému multilokusovému genotypu fenotyp. Ovšem tato metoda opomíjí kvantifikaci nejis-



Obrázek 5.3: Frekvence alel SNP rs16891982 u světových populací

Zdroj: <http://hgdp.uchicago.edu/>

toty, která je nezbytná pro interpretaci výsledků ve forenzní genetice, například při určování fenotypu neznámého vzorku. Proto byla následně použita metoda, která dovolu je určit stupeň nejistoty. Tvorba Bayesových sítí je v tomto případě velmi vhodná, protože dovolu je kombinovat výsledky více genetických analýz, ale zároveň umožňuje kombinaci více důkazů a započtení prvotních předpokladů. Výběr výchozích pravděpodobností pro některé uzly jako „Barva očí“ nebo „Původ“ je nejspornější bod tvorby BS, protože záleží na dostupné podpůrné informaci. Ovšem pokud se tato informace změní, například na základě nových výzkumných studií, lze jednoduše změnit i výchozí pravděpodobnost. Spojení modelů pro určování barvy očí a biogeografického původu zlepšilo výsledky analýzy především u asijských vzorků, protože klesl počet vzorků, u nichž nebylo možné určit barvu očí. Senzitivita určování světlé barvy očí dosáhla pomocí Modelu 2 (rs7495174, rs12913832, rs916977, rs1426654, rs16891982) hodnoty 0,96, což je více než ve studii Pospiech et al. (0,85) [130] a Kastelic et al. (0,86) [72]. Pravděpodobně proto, že součástí Modelu 2 byly markery určující biogeografický původ. Senzitivita sady markerů Irisplex byla vyšší než v této studii - 0,99, i když senzitivita modifikace Irisplex zahrnující rs1129038 byla

stejná jako v této studii - 0,96 [69]. Hodnoty specifcity pro určování světlé barvy očí jsou nižší v této studii (0,98) ve srovnání se studií Pospiech et al. (1,0) [130], Kastelic et al. (0,86) [72] a modifikované sady Irisplex (0,99) [69]. Ovšem senzitivita sady Irisplex byla nižší než v této studii - 0,86 [69]. Důvodem nižší specifcity v této studii je relativně vysoká senzitivita. Senzitivita a specifcita jsou vzájemně spojené, proto vede zvýšená senzitivita ke snížení specifcity a obráceně.

Velmi zajímavý je také rozdíl mezi asijskými vzorky, které byly použity v trénovací a testovací sadě. Jako trénovací byly použity vzorky Kazachů, které jsou, jak se zdá, mnohem méně genotypově jednotné, než populační vzorky Japonců a Číňanů. Jako příklad lze uvést markery rs12913832 a rs1426654, pro které byli jak Japonci, tak i Číňané homozygoté AA a GG. U kazašských vzorků bylo pro marker rs12913832 45% homozygotů AA, 48% heterozygotů AG a 7% homozygotů GG. V případě markeru rs1426654 byl poměr genotypů následující: 35 % GG, 42% AG a 23% AA. Toto genotypové rozdělení podporuje hypotézu, že Kazaši patří ke smíšeným populacím, což bylo zjištěno i v jiných studiích [131]. Napovídá tomu i skutečnost, že území dnešního Kazachstanu dříve protínala hedvábná stezka spojující Evropu a Asii.

6 ZÁVĚR

Studie zkoumající sadu markerů DIPplex obsahující 30 inzerčně-delečních polymorfizmů, které byly umístěny ve vzdálenosti nejméně 10 Mb od všech komerčně dostupných markerů STR. Po optimalizaci pufru může být tato sada použita pro analýzu inhibovaných vzorků. Mezi zkoumanými markery nebyla nalezena statisticky významná nerovnováha, proto mohou být tyto markery považovány za nezávislé. Tyto nové markery mohou být s výhodou analyzovány pomocí metod a přístrojů, které jsou běžně dostupné ve forenzní laboratoři. Díky relativně nízké mutační rychlosti jsou tyto markery velmi vhodné pro příbuzenskou analýzu, obzvláště pokud je podezření na výskyt mutace. Zjištěné hodnoty paternitního indexu ukazují, že sada DIPplex vede k několikanásobnému zvýšení paternitního indexu vytvořeného pomocí analýzy markerů STR. Sada DIPplex může být využita jak při forenzní identifikaci, tak i při testování příbuznosti. Tato studie prokázala, že sada DIPplex je robustní a že je její použití jako doplňkové sady markerů vhodné v České republice.

Studie zaměřená na sadu obsahující 8 vázaných markerů X-STR Argus X-8 prokázala vhodnost použití této sady pro forenzní účely ve středochorvatské populaci. Tyto markery se nacházejí ve 4 vazebných skupinách a jsou tedy velmi užitečné pro příbuzenskou analýzu, i když může v některých případech nastat problém nulových frekvencí zkoumaných haplotypů. Proto jsou tyto markery velmi vhodné pro analýzu příbuznosti kosterních vzorků, kde může jejich použití být více informativní než použití autozomálních markerů STR.

Studie zabývající se sadou Decaplex X-STR představuje návrh sady markerů, který je odlišný od předchozí sady Argus X-8. U všech párů markerů, vyjma páru DXS6809 a DXS6789, mezi kterými je vzdálenost 500 Kb, nebyla zjištěna vazebná nerovnováha. Proto si tyto markery zachovávají vlastnosti spojené s jejich lokalizací na gonozómu X, ale jejich vazebná nezávislost snižuje početní náročnost v případě výpočtu příbuznosti. Výpočet průměrné síly vyloučení poukázal na to, že je tato sada velmi vhodná pro komplexní a deficientní případy určování otcovství a sourozenectví. Protože se výsledky forenzních parametrů markerů obsažených v této sadě shodovaly s jinými studiemi, lze tuto sadu považovat za robustní a za vhodnou pro použití ve forenzní genetice a příbuzenské analýze v České republice.

Hlavním cílem studie fenotypových markerů SNP byla selekce nejlepšího modelu

pro predikci barvy očí a s ní spojeného původu pro podmínky České republiky. Toto je nová oblast forenzní genetiky, může pomoci zúžit okruh možných podezřelých osob či nepřímo podpořit výpověď očitého svědka. Rekonstrukce haplotypů zkoumaných osob podpořila hypotézu, že barva očí se dědí jako recesivní znak. Porovnání výsledků studie s výsledky ostatních studií naznačilo, že existují markery, které mají velkou informativitu bez ohledu na populaci, ve které jsou zkoumány (například rs12913832). Existují také markery, které mají velkou informativitu ohledně predikce barvy očí, ale jenom v určitých populacích. Pomocí analýzy MDR bylo vybráno několik markerů, které nejlépe predikují barvu očí nebo původ. Z těchto výsledků byly následně vytvořeny 2 modely Bayesovských sítí, které byly následně validovány pomocí dalšího souborů vzorků. Kombinace analýzy MDR a Bayesovských sítí se ukázala jako elegantní řešení, které dovoluje zahrnout nejistotu do výpočtu predikce barvy očí a původu. Spojení fenotypových markerů a markerů biogeografického původu vedlo k vyšší míře predikce a nižšímu počtu analýz, u nichž nebylo možné udělat závěr. Doufáme, že poznatky zjištěné pomocí této studie podpoří zavedení metody molekulární fenotypizace v oblasti forenzní genetiky v České republice.

SEZNAM POUŽITÉ LITERATURY

- [1] AJ Jeffreys, V Wilson, SL Thein, et al. Hypervariable minisatellite regions in human DNA. *Nature*, 314(6006):67–73, 1985.
- [2] AJ Jeffreys, JFY Brookfield, and R Semeonoff. Positive identification of an immigration test-case using human DNA fingerprints. *Nature*, 317(6040):818–819, 1985.
- [3] JM Butler. *Fundamentals of forensic DNA typing*. Academic Press, 2009.
- [4] Z Wong, V Wilson, I Patel, S Povey, and AJ Jeffreys. Characterization of a panel of highly variable minisatellites cloned from human DNA. *Ann Hum Genet*, 51(4):269–288, 1987.
- [5] RK Saiki, S Scharf, F Faloona, KB Mullis, GT Horn, HA Erlich, and N Arnheim. Enzymatic amplification of b-globin genomic sequences and restriction

- site analysis for diagnosis of sickle cell anemia. *Science*, 230(4732):1350–1354, 1985.
- [6] R Helmuth, N Fildes, E Blake, MC Luce, J Chimera, R Madej, C Gorodezky, M Stoneking, N Schmill, W Klitz, et al. HLA-DQ α allele and genotype frequencies in various human populations determined by using enzymatic amplification and oligonucleotide probes. *Am J Hum Genet*, 47(3):515, 1990.
 - [7] E Hagelberg, IC Gray, and AJ Jeffreys. Identification of the skeletal remains of a murder victim by DNA analysis. *Nature*, 352(6334):427–429, 1991.
 - [8] CP Kimpton, P Gill, A Walton, A Urquhart, ES Millican, and M Adams. Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *Genome Res*, 3(1):13–22, 1993.
 - [9] TM Clayton, JP Whitaker, and CN Maguire. Identification of bodies from the scene of a mass disaster using DNA amplification of short tandem repeat (STR) loci. *Forensic Sci Int*, 76(1):7–15, 1995.
 - [10] R Hopgood, KM Sullivan, P Gill, et al. Strategies for automated sequencing of human mitochondrial DNA directly from PCR products. *Biotechniques*, 13(1):82, 1992.
 - [11] L Roewer and JT Epplen. Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work. *Forensic Sci Int*, 53(2):163–171, 1992.
 - [12] B Budowle, AJ Eisenberg, and A van Daal. Validity of low copy number typing and applications to forensic science. *Croat Med J*, 50(3):207–217, 2009.
 - [13] P Gill, PL Ivanov, C Kimpton, R Piercy, N Benson, G Tully, I Evett, E Hagelberg, and K Sullivan. Identification of the remains of the Romanov family by DNA analysis. *Nature Genet*, 6(2):130–135, 1994.
 - [14] MD Coble, OM Loreille, MJ Wadhams, SM Edson, K Maynard, CE Meyer, H Niederstätter, C Berger, B Berger, AB Falsetti, et al. Mystery solved: the identification of the two missing Romanov children using DNA analysis. *PLoS One*, 4(3):e4838, 2009.

- [15] MA Jobling and P Gill. Encoded evidence: Dna in forensic analysis. *Nat Rev Genet*, 5(10):739–751, 2004.
- [16] L Roewer, M Krawczak, S Willuweit, M Nagy, C Alves, A Amorim, K Anslinger, C Augustin, A Betz, E Bosch, et al. Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Sci Int*, 118(2):106–113, 2001.
- [17] AT Chinwalla, LL Cook, KD Delehaunty, GA Fewell, LA Fulton, RS Fulton, TA Graves, LW Hillier, ER Mardis, JD McPherson, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915):520–562, 2002.
- [18] PFR Little. Structure and function of the human genome. *Genome Res*, 15(12):1759–1766, 2005.
- [19] G Barbujani, A Magagni, E Minch, and LL Cavalli-Sforza. An apportionment of human dna diversity. *Proc Natl Acad Sci*, 94(9):4516–4519, 1997.
- [20] A Jasinska and WJ Krzyzosiak. Repetitive sequences that shape the human transcriptome. *FEBS letters*, 567(1):136–141, 2004.
- [21] JM Butler. *Advanced Topics in Forensic DNA Typing: Methodology: Methodology*. Academic Press, 2011.
- [22] C Gunter. Genome biology: She moves in mysterious ways. *Nature*, 434(7031):279–280, 2005.
- [23] MT Ross, DV Grafham, AJ Coffey, S Scherer, K McLay, D Muzny, M Platzer, GR Howell, C Burrows, CP Bird, et al. The DNA sequence of the human X chromosome. *Nature*, 434(7031):325–337, 2005.
- [24] MF Lyon. Sex chromatin and gene action in the mammalian X-chromosome. *Am J Hum Genet*, 14(2):135, 1962.
- [25] R Szibor. X-chromosomal markers: past, present and future. *Forensic Sci Int Genet*, 1(2):93–99, 2007.

- [26] R Szibor, M Krawczak, S Hering, J Edelmann, E Kuhlisch, and D Krause. Use of X-linked markers for forensic purposes. *Int J Leg Med*, 117(2):67–74, 2003.
- [27] AL Edwards, HA Hammond, L Jin, CT Caskey, and R Chakraborty. Genetic variation at five trimeric and tetrameric tandem repeat loci in four human population groups. *Genomics*, 12(2):241–253, 1992.
- [28] AR La Spada, EM Wilson, DB Lubahn, AE Harding, and KH Fischbeck. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature*, 352(6330):77–79, 1991.
- [29] R Szibor, S Hering, and J Edelmann. The HumARA genotype is linked to spinal and bulbar muscular dystrophy and some further disease risks and should no longer be used as a DNA marker for forensic purposes. *Int J Leg Med*, 119(3):179–180, 2005.
- [30] L Gusmão, P Sánchez-Diz, C Alves, I Gomes, MT Zarrabeitia, M Abovich, I Atmetlla, C Bobillo, L Bravo, J Builes, et al. A GEP-ISFG collaborative study on the optimization of an X-STR decaplex: data on 15 Iberian and Latin American populations. *Int J Leg Med*, 123(3):227–234, 2009.
- [31] J Schmidtke, W Kühnau, D Wand, J Edelmann, R Szibor, and M Krawczak. Prenatal exclusion without involving the putative fathers of an incestuous father–daughter parenthood. *Prenatal diagnosis*, 24(8):662–664, 2004.
- [32] B Sobrino, M Brión, and A Carracedo. SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci Int*, 154(2):181–194, 2005.
- [33] B Budowle and A van Daal. Forensically relevant SNP classes. *BioTechniques*, 44(5):603–610, 2008.
- [34] JJ Sanchez, C Phillips, C Børsting, K Balogh, M Bogus, M Fondevila, CD Harrison, E Musgrave-Brown, A Salas, PM Schneider, et al. A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis*, 27(9):1713–1724, 2006.

- [35] E Musgrave-Brown, D Ballard, K Balogh, K Bender, B Berger, M Bogus, C Børsting, M Brion, M Fondevila, C Harrison, et al. Forensic validation of the SNPforID 52-plex assay. *Forensic Sci Int Genet*, 1(2):186, 2007.
- [36] L Porras, C Phillips, M Fondevila, L Beltran, T Ortiz, F Rondon, G Barreto, MV Lareu, J Henao, and A Carracedo. Genetic variability of the SNP for ID 52-plex identification-SNP panel in Central West Colombia. *Forensic Sci Int Genet*, 4(1):e9–e10, 2009.
- [37] R Pereira, M Fondevila, C Phillips, A Amorim, A Carracedo, and L Gusmao. Genetic characterization of 52 autosomal SNPs in the Portuguese population. *Forensic Sci Int Genet*, 1(1):358–360, 2008.
- [38] K Drobnič, C Børsting, E Rockenbauer, C Tomas, and N Morling. Typing of 49 autosomal SNPs by SNaPshot in the Slovenian population. *Forensic Sci Int Genet*, 4(5):e125–e127, 2010.
- [39] KK Kidd, AJ Pakstis, WC Speed, EL Grigorenko, SLB Kajuna, NJ Karoma, S Kungulilo, JJ Kim, RB Lu, A Odunsi, et al. Developing a SNP panel for forensic identification of individuals. *Forensic Sci Int*, 164(1):20–32, 2006.
- [40] AA Westen, AS Matai, JFJ Laros, HC Meiland, M Jasper, WJF de Leeuw, P de Knijff, and T Sijen. Tri-allelic SNP markers enable analysis of mixed and degraded DNA samples. *Forensic Sci Int Genet*, 3(4):233–241, 2009.
- [41] J Ge, B Budowle, JV Planz, and R Chakraborty. Haplotype block: a new type of forensic DNA markers. *Int J Leg Med*, 124(5):353–361, 2010.
- [42] T Frudakis. *Molecular photofitting: predicting ancestry and phenotype using DNA*. Academic Press, 2010.
- [43] RE Mills, CT Luttig, CE Larkins, A Beauchamp, C Tsui, WS Pittard, and SE Devine. An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res*, 16(9):1182–1190, 2006.
- [44] JL Weber, D David, J Heil, Y Fan, C Zhao, and G Marth. Human diallelic insertion/deletion polymorphisms. *Am J Hum Genet*, 71(4):854–862, 2002.

- [45] RE Mills, EA Bennett, RC Iskow, and SE Devine. Which transposable elements are active in the human genome? *Trends Genet*, 23(4):183–191, 2007.
- [46] JR Pimenta and SDJ Pena. Efficient human paternity testing with a panel of 40 short insertion-deletion polymorphisms. *Genet Mol Res*, 9(1):601–607, 2010.
- [47] R Pereira, C Phillips, C Alves, A Amorim, A Carracedo, and L Gusmão. A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis*, 30(21):3682–3690, 2009.
- [48] C Li, S Zhao, S Zhang, L Li, Y Liu, J Chen, and J Xue. Genetic polymorphism of 29 highly informative InDel markers for forensic use in the Chinese Han population. *Forensic Sci Int Genet*, 5(1):e27–e30, 2011.
- [49] RA Sturm and M Larsson. Genetics of human iris colour and patterns. *Pigment Cell Melanoma Res*, 22(5):544–562, 2009.
- [50] BA Gilchrest and G Szabo. Effects of aging and chronic sun exposure on melanocytes in human skin. *J Invest Dermatol*, 73(2):141–143, 1979.
- [51] D Scherer and R Kumar. Genetics of pigmentation in skin cancer—a review. *Mutat Res*, 705(2):141, 2010.
- [52] Chi, JC Valencia, ZZ Hu, H Watabe, H Yamaguchi, NJ Mangini, H Huang, VA Canfield, KC Cheng, F Yang, et al. Proteomic and bioinformatic characterization of the biogenesis and function of melanosomes. *J Prot Res*, 5(11):3135–3144, 2006.
- [53] JC García-Borrón, BL Sánchez-Laorden, and C Jiménez-Cervantes. Melanocortin-1 receptor structure and functional regulation. *Pigm Cell Res*, 18(6):393–410, 2005.
- [54] RA Sturm. A golden age of human pigmentation genetics. *Trends Genet*, 22(9):464–468, 2006.
- [55] C Levy, M Khaled, and DE Fisher. MITF: master regulator of melanocyte development and melanoma oncogene. *Trends Mol Med*, 12(9):406–414, 2006.

- [56] PPM Schnetkamp. The SLC24 Na⁺/Ca²⁺-K⁺ exchanger family: vision and beyond. *Pflügers Arch*, 447(5):683–688, 2004.
- [57] J Mengel-From, TH Wong, N Morling, JL Rees, and IJ Jackson. Genetic determinants of hair and eye colours in the Scottish and Danish populations. *BMC Genet*, 10(1):88, 2009.
- [58] J Han, P Kraft, Q Nan, Hand Guo, C Chen, A Qureshi, SE Hankinson, FB Hu, DL Duffy, ZZ Zhao, et al. A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet*, 4(5):e1000074, 2008.
- [59] P Sulem, DF Gudbjartsson, N Stacey, A Helgason, T Rafnar, KP Magnusson, A Manolescu, A Karason, A Palsson, G Thorleifsson, et al. Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nature Genet*, 39(12):1443–1452, 2007.
- [60] H Eiberg, J Troelsen, M Nielsen, A Mikkelsen, J Mengel-From, KW Kjaer, and L Hansen. Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the HERC2 gene inhibiting OCA2 expression. *Hum Genet*, 123(2):177–187, 2008.
- [61] N Goel, M Terman, and J Terman. Depressive symptomatology differentiates subgroups of patients with seasonal affective disorder. *Depress Anxiety*, 15(1):34–41, 2002.
- [62] M Kayser, F Liu, ACJW Janssens, F Rivadeneira, O Lao, K van Duijn, M Vermeulen, P Arp, MM Jhamai, WFJ van IJcken, et al. Three genome-wide association studies and a linkage analysis identify HERC2 as a human iris color gene. *Am J Hum Genet*, 82(2):411, 2008.
- [63] F Liu, A Wollstein, PG Hysi, GA Ankra-Badu, TD Spector, D Park, G Zhu, M Larsson, DL Duffy, GW Montgomery, et al. Digital quantification of human eye color highlights genetic association of three new loci. *PLoS Genet*, 6(5):e1000934, 2010.
- [64] P Sulem, D Gudbjartsson, SN Stacey, A Helgason, T Rafnar, M Jakobsdottir, S Steinberg, SA Gudjonsson, A Palsson, G Thorleifsson, et al. Two newly

- identified genetic determinants of pigmentation in Europeans. *Nature Genet*, 40(7):835–837, 2008.
- [65] C Bouakaze, C Keyser, E Crubézy, D Montagnon, and B Ludes. Pigment phenotype and biogeographical ancestry from ancient skeletal remains: inferences from multiplexed autosomal SNP analysis. *Int J Leg Med*, 123(4):315–325, 2009.
- [66] S Walsh, F Liu, KN Ballantyne, M van Oven, O Lao, and M Kayser. IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Science International: Genetics*, 5(3):170–180, 2011.
- [67] S Walsh, A Wollstein, F Liu, U Chakravarthy, M Rahu, JH Seland, G Soubrane, L Tomazzoli, F Topouzis, JR Vingerling, et al. DNA-based eye colour prediction across Europe with the IrisPlex system. *Forensic Sci Int Genet*, 6(3):330–340, 2012.
- [68] NA Rosenberg. Standardized subsets of the HGDP-CEPH Human Genome Diversity Cell Line Panel, accounting for atypical and duplicated samples and pairs of close relatives. *Ann Hum Genet*, 70(6):841–847, 2006.
- [69] Y Ruiz, C Phillips, A Gomez-Tato, J Alvarez-Dios, M Casares de Cal, R Cruz, O Maroñas, J Söchtig, M Fondevila, MJ Rodriguez-Cid, et al. Further development of forensic eye color predictive tests. *Forensic Sci Int Genet*, 7(1):28–40, 2013.
- [70] O Spichenok, ZM Budimlija, AA Mitchell, A Jenny, L Kovacevic, D Marjanovic, T Caragine, M Prinz, and E Wurmbach. Prediction of eye and skin color in diverse populations using seven SNPs. *Forensic Sci Int Genet*, 5(5):472–478, 2011.
- [71] E Pośpiech, J Draus-Barini, T Kupiec, A Wojas-Pelc, and W Branicki. Gene–gene interactions contribute to eye colour variation in humans. *J Hum Genet*, 56(6):447–455, 2011.
- [72] V Kastelic and K Drobnič. A single-nucleotide polymorphism (SNP) multiplex system: the association of five SNPs with human eye and hair color in the

- Slovenian population and comparison using a Bayesian network and logistic regression model. *Croat Med J*, 53(5):401–408, 2012.
- [73] DM Altshuler, RA Gibbs, L Peltonen, E Dermitzakis, SF Schaffner, F Yu, PE Bonnen, PI De Bakker, P Deloukas, SB Gabriel, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*, 467(7311):52, 2010.
 - [74] R Ihaka and R Gentleman. R: A language for data analysis and graphics. *J Comput Graph Stat*, 5(3):299–314, 1996.
 - [75] JS Velarde-Félix, J Salazar-Flores, G Martínez-Cortés, A Flores García, JF Muñoz-Valle, JJ Ríos-Tostado, R Rubi-Castellanos, and H Rangel-Villalobos. Deletion mapping and paternal origin of a Mexican AMELY negative male. *Leg Med*, 13(5):262–264, 2011.
 - [76] K Thangaraj, AG Reddy, and L Singh. Is the amelogenin gene reliable for gender identification in forensic casework and prenatal diagnosis? *Int J Leg Med*, 116(2):121–123, 2002.
 - [77] X Ou, W Chen, H Chen, F Zhao, J Zheng, D Tong, Y Chen, A Chen, and H Sun. Null alleles of the X and Y chromosomal amelogenin gene in a Chinese population. *Int J Leg Med*, 126(4):513–518, 2012.
 - [78] R Kumagai, Y Sasaki, T Tokuta, H Biwasaka, A Matsusue, Y Aoki, and K Dewa. Distinct breakpoints in two cases with deletion in the Yp11. 2 region in Japanese population. *Hum Genet*, 127(5):537–543, 2010.
 - [79] S Turrina, G Filippini, G Voglino, and D De Leo. Two additional reports of deletion on the short arm of the Y chromosome. *Forensic Sci Int Genet*, 5(3):242–246, 2011.
 - [80] S Myles, M Somel, K Tang, J Kelso, and M Stoneking. Identifying genes underlying skin pigmentation differences among human populations. *Human Genet*, 120(5):613–621, 2007.
 - [81] W Branicki, U Brudnik, T Kupiec, P Wolańska-Nowak, A Szczerbińska, and A Wojas-Pelc. Association of polymorphic sites in the OCA2 gene with eye colour using the tree scanning method. *Ann Hum Genet*, 72(2):184–192, 2007.

- [82] I Yuasa, K Umetsu, S Harihara, A Kido, A Miyoshi, N Saitou, B Dashnyam, F Jin, G Lucotte, PK Chattopadhyay, and other. Distribution of two Asian-related coding SNPs in the MC1R and OCA2 genes. *Biochem Genet*, 45(7):535–542, 2007.
- [83] RA Sturm, DL Duffy, ZZ Zhao, FPN Leite, MS Stark, NK Hayward, NG Martin, and GW Montgomery. A single SNP in an evolutionary conserved region within intron 86 of the HERC2 gene determines human blue-brown eye color. *Am J Hum Genet*, 82(2):424, 2008.
- [84] H Nan, P Kraft, AA Qureshi, Q Guo, C Chen, SE Hankinson, FB Hu, G Thomas, RN Hoover, S Chanock, et al. Genome-wide association study of tanning phenotype in a population of European ancestry. *J Invest Dermatol*, 129(9):2250–2257, 2009.
- [85] IV Kutuyavin, IA Afonina, A Mills, VV Gorn, EA Lukhtanov, ES Belousov, MJ Singer, DK Walburger, SG Lokhov, AA Gall, et al. 3 - Minor groove binder-DNA probes increase sequence specificity at PCR extension temperatures. *Nucleic Ac Res*, 28(2):655–661, 2000.
- [86] J Xu, A Turner, J Little, ER Bleecker, and DA Meyers. Positive results in association studies are associated with departure from Hardy-Weinberg equilibrium: hint for genotyping error? *Hum Genet*, 111(6):573–574, 2002.
- [87] L Hosking, S Lumsden, K Lewis, A Yeo, L McCarthy, A Bansal, J Riley, I Purvis, and CF Xu. Detection of genotyping errors by Hardy-Weinberg equilibrium testing. *Eur J Hum Genet*, 12(5):395–399, 2004.
- [88] DM Nielsen, MG Ehm, BS Weir, et al. Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. *Am J Hum Genet*, 63:1531–1540, 1998.
- [89] R Chakraborty, PE Smouse, and JV Neel. Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. *Am J Hum Genet*, 43(5):709, 1988.

- [90] L Excoffier, G Laval, and S Schneider. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online*, 1:47, 2005.
- [91] S Purcell, B Neale, K Todd-Brown, L Thomas, MAR Ferreira, D Bender, J Maller, P Sklar, PIW De Bakker, MJ Daly, et al. Plink: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81(3):559–575, 2007.
- [92] RJ Cabin and RJ Mitchell. To Bonferroni or not to Bonferroni: when and how are the questions. *Bul Ecol Soc Am*, 81(3):246–248, 2000.
- [93] D Botstein, RL White, M Skolnick, and RW Davis. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet*, 32(3):314, 1980.
- [94] RA Fisher. Standard calculations for evaluating a blood-group system. *Heredity*, 5:95–102, 1951.
- [95] D Desmarais, Y Zhong, R Chakraborty, C Perreault, L Busque, et al. Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA). *J Forensic Sci*, 43(5):1046–1049, 1998.
- [96] J Krüger, W Fuhrmann, K-H Lichte, and Chr Steffens. Zur Verwendung des Polymorphismus der sauren Erythrocytenphosphatase bei der Vaterschaftsbeurteilung. *Int Journal Leg Med*, 64(2):127–146, 1968.
- [97] CH Brenner. Symbolic kinship program. *Genetics*, 145(2):535–542, 1997.
- [98] M Krawczak. Informativity assessment for biallelic single nucleotide polymorphisms. *Electrophoresis*, 20(8):1676–1681, 1999.
- [99] S Wright. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, pages 395–420, 1965.
- [100] L Excoffier, G Laval, and D Balding. Gametic phase estimation over large genomic regions using an adaptive window approach. *Hum Genomics*, 1(1):7–19, 2003.

- [101] JM Bland and DG Altman. Statistics notes: the odds ratio. *Br Med J*, 320(7247):1468, 2000.
- [102] HTO Davies and IK Crombie. What are confidence intervals and p-values. *London: Hayward Medical Communications*, 2009.
- [103] CM Lewis and J Knight. Introduction to Genetic Association Studies. *Cold Spring Harb Protoc*, 2012(3):297–306, 2012.
- [104] MD Ritchie, LW Hahn, and JH Moore. Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. *Genet Epidemiol*, 24(2):150–157, 2003.
- [105] J Gui, AS Andrew, P Andrews, HM Nelson, KT Kelsey, MR Karagas, and JH Moore. A robust multifactor dimensionality reduction method for detecting gene-gene interactions with application to the genetic analysis of bladder cancer susceptibility. *Ann Hum Genet*, 75(1):20–28, 2011.
- [106] JH Moore. Computational analysis of gene-gene interactions using multifactor dimensionality reduction. *Expert Rev Mol Diagn*, 4(6):795–803, 2004.
- [107] M Robnik-Šikonja and I Kononenko. Theoretical and empirical analysis of ReliefF and RReliefF. *Mach Learn*, 53(1):23–69, 2003.
- [108] AA Motsinger and MD Ritchie. Multifactor dimensionality reduction: An analysis strategy for modelling and detecting gene-gene interactions in human genetics and pharmacogenomics studies. *Hum Genomics*, 2(5):318–328, 2006.
- [109] D Becker, H Rodig, C Augustin, J Edelmann, F Götz, S Hering, R Szibor, and W Brabetz. Population genetic evaluation of eight X-chromosomal short tandem repeat loci using Mentype Argus X-8 PCR amplification kit. *Forensic Sci Int Genet*, 2(1):69–74, 2008.
- [110] W Branicki, P Wolańska-Nowak, A Parys-Proszek, and T Kupiec. Application of the Mentype Argus X-8 kit to forensic casework. *Prob Forensic Sci*, 73:53–64, 2008.

- [111] TM Diegoli, L Kovacevic, N Pojskic, MD Coble, and D Marjanovic. Population study of fourteen X chromosomal short tandem repeat loci in a population from Bosnia and Herzegovina. *Forensic Sci Int Genet*, 5(4):350–351, 2011.
- [112] I Gomes, V Pereira, V Gomes, MJ Prata, N Pinto, A Carracedo, A Amorim, and L Gusmão. The Karimojong from Uganda: genetic characterization using an X-STR decaplex system. *Forensic Sci Int Genet*, 3(4):e127–e128, 2009.
- [113] B García, M Crespillo, M Paredes, and JL Valverde. Population data for 10 X-chromosome STRs from north-east of Spain. *Forensic Sci Int Genet*, 6(1):e13, 2012.
- [114] M Poetsch, H Petersmann, A Repenning, and E Lignitz. Development of two pentaplex systems with X-chromosomal STR loci and their allele frequencies in a northeast german population. *Forensic Sci Int*, 155(1):71–76, 2005.
- [115] W Pepinski, M Skawronska, A Niemcunowicz-Janica, E Koc-Zorawska, J Janica, and I Soltyszewski. Polymorphism of four X-chromosomal STRs in a Polish population sample. *Forensic Sci Int*, 151(1):93–95, 2005.
- [116] SL Friis, C Børsting, E Rockenbauer, L Poulsen, SF Fredslund, C Tomas, and N Morling. Typing of 30 insertion/deletions in Danes using the first commercial indel kit—Mentype[®] DIPplex. *Forensic Sci Int Genet*, 6(2):e72–e74, 2012.
- [117] P Martín, O García, B Heinrichs, I Yurrebaso, A Aguirre, and A Alonso. Population genetic data of 30 autosomal indels in Central Spain and the Basque Country populations. *Forensic Sci Int Genet*, 7(1):e27–e30, 2012.
- [118] AM Neuvonen, JU Palo, M Hedman, and A Sajantila. Discrimination power of Investigator DIPplex loci in Finnish and Somali populations. *Forensic Sci Int Genet*, 6(4):e99–e102, 2012.
- [119] EH Kim, HY Lee, IS Yang, WI Yang, and KJ Shin. Population data for 30 insertion–deletion markers in a Korean population. *Int J Leg Med*, pages 1–2, 2013.

- [120] A Zidkova, A Horinek, V Kebrdlova, and M Korabecna. Application of the new insertion–deletion polymorphism kit for forensic identification and parentage testing on the Czech population. *Int J Leg Med*, 127(1):7–10, 2013.
- [121] W Bär, B Brinkmann, B Budowle, A Carracedo, P Gill, P Lincoln, W Mayr, and B Olaisen. DNA recommendations. *Int J Leg Med*, 110(4):175–176, 1997.
- [122] M Gelabert-Besada, S Ferreira, M García-Magariños, L Gusmão, and P Sánchez-Diz. Genetic characterization of Western Iberia using Mentype® Argus X-8 kit. *Forensic Sci Int Genet*, 6(1):e39, 2012.
- [123] H Luo, Y Ye, Y Wang, W Liang, L Yun, M Liao, J Yan, J Wu, Y Li, and Y Hou. Characteristics of eight X-STR loci for forensic purposes in the Chinese population. *Int J Leg Med*, 125(1):127–131, 2011.
- [124] N Cerri, A Verzeletti, F Gasparini, A Poglio, E Mazzeo, and F De Ferrari. Population data for 8 X-chromosome STR loci in a population sample from Northern Italy and from the Sardinia island. *Forensic Sci Int Genet Supplement Series*, 1(1):173–175, 2008.
- [125] A Zidkova, P Coufalova, and P Capek. X-STR decaplex study on the population of Czech Republic. *Int J Leg Med*, pages 1–2, 2013.
- [126] M Soejima and Y Koda. Population differences of two coding SNPs in pigmentation-related genes SLC24A5 and SLC45A2. *Int J Leg Med*, 121(1):36–39, 2007.
- [127] BK Rana, D Hewett-Emmett, L Jin, BHJ Chang, N Sambuughin, M Lin, S Watkins, M Bamshad, LB Jorde, M Ramsay, et al. High polymorphism at the human melanocortin 1 receptor locus. *Genetics*, 151(4):1547–1557, 1999.
- [128] R Iida, M Ueki, H Takeshita, J Fujihara, T Nakajima, Y Kominato, M Nagao, and T Yasuda. Genotyping of five single nucleotide polymorphisms in the OCA2 and HERC2 genes associated with blue-brown eye color in the Japanese population. *Cell Biochem Funct*, 27(5):323–327, 2009.
- [129] DL Duffy, GW Montgomery, W Chen, ZZ Zhao, L Le, MR James, NK Hayward, NG Martin, and RA Sturm. A three–single-nucleotide polymorphism

haplotype in intron 1 of OCA2 explains most human eye-color variation. *Am J Hum Genet*, 80(2):241, 2007.

- [130] E Pośpiech, J Draus-Barini, T Kupiec, A Wojas-Pelc, and W Branicki. Prediction of eye color from genetic data using bayesian approach. *J For Sci*, 57(4):880–886, 2012.
- [131] B Keating, AT Bansal, S Walsh, J Millman, J Newman, K Kidd, B Budowle, A Eisenberg, J Donfack, P Gasparini, et al. First all-in-one diagnostic tool for DNA intelligence: genome-wide inference of biogeographic ancestry, appearance, relatedness, and sex with the Identitas v1 Forensic Chip. *Int J Leg Med*, pages 1–14, 2012.